DOCUMENT RESUME

ED 404 377 TM 026 463

AUTHOR	Allen, Nancy L.; And Others
TITLE	The NAEP 1994 Technical Report.
INSTITUTION	Educational Testing Service, Princeton, N.J.; National Assessment of Educational Progress, Princeton, NJ.
SPONS AGENCY	National Center for Education Statistics (ED), Washington, DC.
REPORT NO	NCES-97-897
PUB DATE	Dec 96
NOTE	1,032p.
PUB TYPE	Reports - Evaluative/Feasibility (142)
EDRS PRICE	MF07/PC42 Plus Postage.
DESCRIPTORS	*Academic Achievement; *Elementary Secondary Education; National Surveys; Research Design; *Research Methodology; Research Reports
IDENTIFIERS	*National Assessment of Educational Progress

ABSTRACT

The 1994 National Assessment of Educational Progress (NAEP) monitored the performance of students in American schools in reading, U.S. history, geography, mathematics, science, and writing. The sample involved more than 135,000 public and private school students who were 9, 13, or 17 years old and/or in grades 4, 8, and 12. This technical report provides details on instrument development, sample design, data collection, and data analysis procedures for the 1994 NAEP. Steps involved in the NAEP are described from planning through the creation of edited data files. Detailed substantive results are presented in a series of reports by the National Center for Education Statistics. Part I of this report presents details of the design of the assessment and various data collection procedures with chapters on the individual assessments. Part II describes the data analysis procedures and discusses the weighting and variance estimation procedures, the scaling methodology, and the trend and main assessment analyses overall and for the subjects assessed. Chapter 19 presents the basic data from the 1994 assessment. Eleven appendixes contain supplemental information with specific information about aspects of the research methodology and data interpretation. (Contains 269 tables and 29 figures.) (SLD)

****
*   Reproductions supplied by EDRS are the best that can be made   *
*   from the original document.   *
****

1M

ED 404 377

TM026463

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

THE NATION'S REPORT CARD naep

Prepared by Educational Testing Service under contract with the National Center for Education Statistics

Office of Educational Research and Improvement
U.S. Department of Education

BEST COPY AVAILABLE BES

What is The Nation's Report Card?

THE NATION'S REPORT CARD, the National Assessment of Educational Progress (NAEP), is the only nationally representative and continuing assessment of what America's students know and can do in various subject areas. Since 1969, assessments have been conducted periodically in reading, mathematics, science, writing, history/geography, and other fields. By making objective information on student performance available to policymakers at the national, state, and local levels, NAEP is an integral part of our nation's evaluation of the condition and progress of education. Only information related to academic achievement is collected under this program. NAEP guarantees the privacy of individual students and their families.

NAEP is a congressionally mandated project of the National Center for Education Statistics, the U.S. Department of Education. The Commissioner of Education Statistics is responsible, by law, for carrying out the NAEP project through competitive awards to qualified organizations. NAEP reports directly to the Commissioner, who is also responsible for providing continuing reviews, including validation studies and solicitation of public comment, on NAEP's conduct and usefulness.

In 1988, Congress established the National Assessment Governing Board (NAGB) to formulate policy guidelines for NAEP. The Board is responsible for selecting the subject areas to be assessed from among those included in the National Education Goals; for setting appropriate student performance levels; for developing assessment objectives and test specifications through a national consensus approach; for designing the assessment methodology; for developing guidelines for reporting and disseminating NAEP results; for developing standards and procedures for interstate, regional, and national comparisons; for determining the appropriateness of test items and ensuring they are free from bias; and for taking actions to improve the form and use of the National Assessment.

The National Assessment Governing Board

Honorable William T. Randall, Chair
Commissioner of Education
State of Colorado
Denver, Colorado

Mary R. Blanton
Attorney
Salisbury, North Carolina

Patsy Cavazos
Principal
W.G. Love Accelerated Elementary School
Houston, Texas

Catherine A. Davidson
Secondary Education Director
Central Kitsap School District
Silverdale, Washington

Edward Donley
Former Chairman
Air Products & Chemicals, Inc.
Allentown, Pennsylvania

Honorable James Edgar
Governor of Illinois
Springfield, Illinois

James E. Ellingson
Fourth-grade Classroom Teacher
Probstfield Elementary School
Moorhead, Minnesota

Thomas H. Fisher
Director, Student Assessment Services
Florida Department of Education
Tallahassee, Florida

Michael J. Guerra
Executive Director
Secondary Schools Department
National Catholic Educational Association
Washington, DC

Jan B. Loveless
Coordinator of Fund Development
Midland Public Schools
Midland, Michigan

Marilyn McConachie
School Board Member
Glenbrook High Schools
Glenview, Illinois

William J. Moloney
Superintendent of Schools
Calvert County Public Schools
Prince Frederick, Maryland

Honorable Annette Morgan
Member
Missouri House of Representatives
Jefferson City, Missouri

Mark D. Musick
President
Southern Regional Education Board
Atlanta, Georgia

Mitsugi Nakashima
President
Hawaii State Board of Education
Honolulu, Hawaii

Michael T. Nettles
Professor of Education & Public Policy
University of Michigan
Ann Arbor, Michigan
and Director
Frederick D. Patterson Research Institute
United Negro College Fund

Honorable Norma Paulus
Superintendent of Public Instruction
Oregon State Department of Education
Salem, Oregon

Honorable Roy Romer
Governor of Colorado
Denver, Colorado

Honorable Edgar D. Ross
Judge
Territorial Court of the Virgin Islands
Christiansted, St. Croix
U.S. Virgin Islands

Fannie L. Simmons
Mathematics Coordinator
District 5 of Lexington/Richland County
Ballentine, South Carolina

Adam Urbanski
President
Rochester Teachers Association
Rochester, New York

Deborah Voltz
Assistant Professor
Department of Special Education
University of Louisville
Louisville, Kentucky

Marilyn A. Whirry
Twelfth-grade English Teacher
Mira Costa High School
Manhattan Beach, California

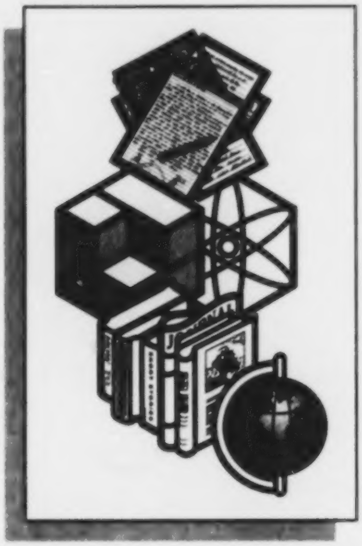
Dennie Palmer Wolf
Senior Research Associate
Harvard Graduate School of Education
Cambridge, Massachusetts

Sharon P. Robinson (Ex-Officio)
Assistant Secretary
Office of Educational Research
and Improvement
U.S. Department of Education
Washington, DC

Roy Truby
Executive Director, NAGB
Washington, DC

NATIONAL CENTER FOR EDUCATION STATISTICS

The NAEP
1994 Technical Report



Nancy L. Allen, Debra L. Kline, and Christine A. Zelenak

in collaboration with

Luz Bay
Patrick B. Bourgeacq
Mary Lyn Bourque
Charles L. Brungardt
Nancy W. Caldwell
Jay R. Campbell
James E. Carlson
Hua-hua Chang
C. Janet Chen
Wen-Hung Chen
Patricia Donahue
John R. Donoghue
John J. Ferris
David S. Freund
Claudia A. Gentile
Edward Ip
Steven P. Isham
Frank Jenkins
Eugene G. Johnson
Susan C. Loomis
Bruce A. Kaplan
Edward Kulick
Stephen Lazer
Phillip T. Leung
John Mazzeo
Robert J. Mislevy
Eiji Muraki
Patricia E. O'Reilly
Katharine E. Pashley
Timothy Robinson
Alfred M. Rogers
Keith F. Rust
Juliet Shaffer
Spencer S. Swinton
Neal Thomas
Ellen Tubbs
Mark M. Waksberg
Leslie Wallace
Minhwei Wang
Steve Wang
Lois H. Worthington

December 1996

THE NATION'S
REPORT
CARD
naep

Prepared by Educational Testing Service under contract with the National Center for Education Statistics

Office of Educational Research and Improvement
U.S. Department of Education

U.S. Department of Education
Richard W. Riley
Secretary

Office of Educational Research and Improvement
Sharon P. Robinson
Assistant Secretary

National Center for Education Statistics
Pascal D. Forgione, Jr., Ph.D.
Commissioner

Education Assessment Group
Gary W. Phillips
Associate Commissioner

FOR MORE INFORMATION:

For ordering information on this report, write:

National Library of Education
Office of Educational Research and Improvement
U.S. Department of Education
555 New Jersey Avenue, NW
Washington, D.C. 20208-5641

or call 1-800-424-1616 (in the Washington, D.C. metropolitan area call 202-219-1651).

The work upon which this publication is based was performed for the National Center for Education Statistics, Office of Educational Research and Improvement, by Educational Testing Service.

Educational Testing Service is an equal opportunity, affirmative action employer.

Educational Testing Service, ETS, and the ETS logo are registered trademarks of Educational Testing Service.

THE NAEP 1994 TECHNICAL REPORT

◆ Table of Contents ◆

List of Tables ... xiv
List of Figures ... xxiii
Acknowledgments ... xxv

Introduction *Juliet Shaffer and James E. Carlson* ... 1

Part I
The Design and Implementation of the 1994 NAEP

Chapter 1 Overview of Part I: The Design and Implementation of the 1994 NAEP
Nancy L. Allen and Eugene G. Johnson ... 5

1.1 Introduction ... 5
1.2 The 1994 NAEP Design ... 6

- 1.2.1 The 1994 NAEP Samples ... 9
- 1.2.2 NAEP Assessments since 1969 ... 11

1.3 Development of Assessment Objectives, Items, and Background Questions ... 11
1.4 The 1994 Sample Design ... 18
1.5 Assessment Instruments ... 21

- 1.5.1 Student Assessment Booklets—Main Assessment ... 21
- 1.5.2 Student Assessment Booklets—Long-Term Trend Samples ... 25
- 1.5.3 Other Instruments ... 26

1.6 Field Operations and Data Collection ... 26
1.7 Materials and Data Processing ... 27
1.8 Professional Scoring ... 28
1.9 Creation of the Database ... 28

Chapter 2 Developing the NAEP Objectives, Items, and Background Questions for the 1994 Assessments of Reading, U.S. History, and Geography
Stephen Lazer, Jay R. Campbell, and Patricia Donahue ... 29

2.1 Introduction ... 29
2.2 General Overview of the 1994 Assessment Frameworks ... 31
2.3 General Overview of Procedures for Developing the Items ... 32
2.4 Developing the Reading Assessment ... 34

2.4.1 Overview 34
2.4.2 Development of the Assessment Framework 34
2.4.3 Framework and Assessment Design Principles 35
2.4.4 Framework for the 1994 Assessment 36
2.4.5 Developing the Cognitive Items 40
2.4.6 Development of the Operational Forms 40
2.4.7 Distribution of Assessment Items 41
2.4.8 Reading Background Questionnaires 43

2.5 Developing the U.S. History Assessment 45

2.5.1 Overview 45
2.5.2 Development of the Assessment Framework 46
2.5.3 Framework for the 1994 Assessment 47
2.5.4 Developing the Cognitive Items 49
2.5.5 Development of the Operational Forms 49
2.5.6 Distribution of Assessment Items 50
2.5.7 U.S. History Background Questionnaires 53
2.5.8 Pilot Study in Group Assessment 55

2.5.8.1 General Description of the Study 55

2.6 Developing the Geography Assessment 56

2.6.1 Overview 56
2.6.2 Development of the Assessment Framework 57
2.6.3 Framework for the 1994 Assessment 58
2.6.4 Developing the Cognitive Items 59
2.6.5 Development of the Operational Forms 60
2.6.6 Distribution of Assessment Items 61
2.6.7 Geography Background Questionnaires 63

2.7 Background Questionnaires 65

2.7.1 Student Questionnaires 65
2.7.2 Teacher, School, and IEP/LEP Student Questionnaires 66

Chapter 3 Sample Design
Leslie Wallace and Keith F. Rust 69

3.1 Introduction 69
3.2 Primary Sampling Units 71
3.3 Selection of Schools 73
3.4 Assignment of Sessions to Schools 78
3.5 Sampling Students 79
3.6 Excluded Students 80
3.7 Student Participation Rates 83

3.8 Overall Student Participation Rates 83
3.9 Sampling Teachers 86

Chapter 4 Assessment Instruments

Stephen Lazer 87

4.1 Introduction 87
4.2 Student Booklets—Main Assessments 87

  4.2.1 Reading 87
  4.2.2 U.S. History 94
  4.2.3 Geography 100

4.3 Student Booklets—Long-Term Trend Assessments 106
4.4 Student, Teacher, and School Questionnaires 112

  4.4.1 Student Questionnaires 112
  4.4.2 IEP/LEP Student Questionnaires 112
  4.4.3 Teacher Questionnaires 113
  4.4.4 School Characteristics and Policies Questionnaire 113

Chapter 5 Field Operations and Data Collection

Nancy W. Caldwell and Mark M. Waksberg 115

5.1 Introduction 115

  5.1.1 Field Organization and Schedule of Activities 115

5.2 Preparing for the Assessments 116

  5.2.1 Gaining the Cooperation of Sampled Schools 116
  5.2.2 Supervisor Training 117
  5.2.3 Recruiting, Hiring, and Training Exercise Administrators 117
  5.2.4 Contacting Districts and Nonpublic Schools 118

5.3 Data Collection for the Main Assessments 119

  5.3.1 Selecting the Student Sample 119
  5.3.2 Conducting the Assessment Sessions 122
  5.3.3 Participation Results for the Main NAEP Assessments 122
  5.3.4 The U.S. History Special Study 124

5.4 Data Collection for the Long-Term Trend Assessments 127

  5.4.1 Selecting the Student Sample 127
  5.4.2 Conduct of the Assessment 128

5.4.3 Participation Results for the Long-Term Trend Assessments 128

5.5 Field Management 131

Chapter 6 Processing Assessment Materials

Patrick B. Bourgeacq and Timothy Robinson 133

6.1 Introduction 133

6.1.1 Innovations for 1994 134

6.2 Printing 141
6.3 Packaging And Shipping 142

6.3.1 Distribution 142

6.4 Processing 143

6.4.1 Overview 143
6.4.2 Document Receipt 144
6.4.3 Batching and Scanning Documents 144
6.4.4 Questionnaires 145
6.4.5 Booklet Accountability 145
6.4.6 Data Entry 146
6.4.7 Data Validation 147
6.4.8 Editing Non-Image and Key-Entered Documents 149
6.4.9 Data Validation and Editing of Image-Processed Documents 150
6.4.10 Data Transmission 150

6.5 Data Delivery 151
6.6 Miscellaneous 153

6.6.1 Storage of Documents 153
6.6.2 Quality Control Documents 153
6.6.3 Alert Analysis 153

Chapter 7 Professional Scoring

Claudia A. Gentile, Charles L. Brungardt, and Patrick B. Bourgeacq 155

7.1 Introduction 155
7.2 Long-Term Trend Assessments 158

7.2.1 Characteristics of the Scoring Guides 158
7.2.2 Training 159
7.2.3 Scoring 160
7.2.4 Holistic Writing Scoring 161

7.2.5 Mechanics Scoring 162

7.3 Main NAEP Assessment - Overview 162

7.3.1 Preparation for Training 162
7.3.2 General Training Guidelines 163
7.3.3 Qualification and Reliability Reports 164
7.3.4 Main U.S. History Assessment 165

7.3.4.1 Characteristics of the Scoring Guides 165
7.3.4.2 Training 166
7.3.4.3 Scoring 166

7.3.5 Main Geography Assessment 167

7.3.5.1 Characteristics of the Scoring Guides 167
7.3.5.2 Training 167
7.3.5.3 Scoring 168

7.3.6 Main Reading Assessment 169

7.3.6.1 Characteristics of the Scoring Guides 169
7.3.6.2 Training 169
7.3.6.3 Scoring 170
7.3.6.4 NAEP Reader 171
7.3.6.5 1992 Short-Term Trend and Image/Paper Special Study 172

7.3.7 Calibration Bridges 173

Chapter 8 Creation of the Database, Quality Control of Data Entry, and Creation of the Database Products
*John J. Ferris, David S. Freund, Patricia E. O'Reilly, Katharine E. Pashley,
and Alfred M. Rogers* 177

8.1 Introduction 177
8.2 Creation of the Database 177

8.2.1 Merging Files 177
8.2.2 Creating the Master Catalog 178

8.3 Quality Control of NAEP Data Entry for 1994 179

8.3.1 Student Data 182
8.3.2 IEP/LEP Student Questionnaire Data 183
8.3.3 Teacher Questionnaire Data 183
8.3.4 School Characteristics and Policies Questionnaire Data 183

8.4 NAEP Database Products 184

8.4.1 The Item Information Database 184
8.4.2 The Secondary-Use Data Files 184

8.4.2.1 File Definition 185
8.4.2.2 Definition of the Variables 185
8.4.2.3 Data Definition 186
8.4.2.4 Data File Layouts 187
8.4.2.5 Data File Catalogs 188
8.4.2.6 Data Codebooks 188
8.4.2.7 Control Statement Files for Statistical Packages 189
8.4.2.8 Machine-Readable Catalog Files 190
8.4.2.9 NAEP Data on Disk 190

Part II
The Analysis of 1994 NAEP Data

Chapter 9 Overview of Part II: The Analysis of 1994 NAEP Data
Nancy L. Allen and James E. Carlson 191

9.1 Introduction 191
9.2 Samples of Students 192
9.3 Analysis Steps 193

9.3.1 Preparation of Final Sampling Weights 193
9.3.2 Analysis of Item Properties: Background and Cognitive Items 193
9.3.3 Scaling 197
9.3.4 Drawing Inferences from the Results 201
9.3.5 Minimum Sample Sizes for Reporting Subgroup Results 202
9.3.6 Estimates of Standard Errors with Large Mean Squared Errors 203
9.3.7 Teacher Questionnaires 203
9.3.8 Dimensionality Analysis 204

9.4 Overview of Chapters 10 through 19 207

Chapter 10 Weighting Procedures and Estimation of Sampling Variance
Eugene G. Johnson, Leslie Wallace, Keith F. Rust, and Ellen Tubbs 209

10.1 Introduction 209
10.2 Derivation of the Sample Weights 211

10.2.1 Student Base Weight 212
10.2.2 Adjustment of Base Weights for Nonresponse 213

10.2.2.1 Session Nonresponse Adjustment (SESNRF) 213
10.2.2.2 Student Nonresponse Adjustment (STUNRF) 214

10.2.2.3 Evaluation of Potential for Bias Resulting from School and Student Nonparticipation 215

10.2.3 Trimming of Weights 224
10.2.4 Poststratification 226

10.2.4.1 1994 Poststratification Procedures 226

10.2.5 The Final Student Weight 227
10.2.6 Other Weights 230

10.3 Procedures Used by NAEP to Estimate Sampling Variability 232
10.4 Approximating the Sampling Variance Using Design Effects 235

Chapter 11 Scaling Procedures
Nancy L. Allen, Eugene G. Johnson, Robert J. Mislevy, and Neal Thomas 247

11.1 Introduction 247
11.2 Background 247
11.3 Scaling Methodology 249

11.3.1 The Scaling Models 249
11.3.2 An Overview of Plausible Values Methodology 256
11.3.3 Computing Plausible Values in IRT-Based Scales 257

11.4 Analyses 259

11.4.1 Computational Procedures 259
11.4.2 Statistical Tests 260
11.4.3 Biases in Secondary Analyses 261
11.4.4 A Numerical Example 262

11.5 Describing Student Performance 263

11.5.1 Achievement Levels 264
11.5.2 Performance Descriptions Based on Composite Scales 264
11.5.3 Item Mapping Procedures 265

11.6 Overview of the 1994 NAEP Scales 266

Chapter 12 Data Analysis for the Reading Assessment
John R. Donoghue, Steven P. Isham, and Lois H. Worthington 267

12.1 Introduction 267
12.2 Description of Student Sample, Items, and Assessment Booklets 268
12.3 Item Analyses 271

12.3.1 Conventional Item and Test Analyses 271
12.3.2 Scoring the Constructed-Response Items 278
12.3.3 Differential Item Functioning 278

12.4 IRT Scaling 281

12.4.1 Overview of Item Parameter Estimation 281
12.4.2 Reanalysis of the 1992 Reading Assessment 283
12.4.3 Evaluation of Model Fit 284

12.4.3.1 Items Deleted from the Final Scale 284

12.4.3.1.1 Recoded Polytomous Items 285

12.4.3.2 Fit Across of a Common Item Category Response Function Across Assessment Years 293

12.4.3.2.1 Short-Term Trend Scorer/Item Parameter Drift 293

12.4.3.2.1.1 Preliminary Screening Based on Rescore Data 294
12.4.3.2.1.2 Identification of Drift Items Based on IRT Model-Fit Plots 295

12.5 Generation of Plausible Values 303
12.6 The Final Reading Scales 305

12.6.1 Purpose-for-Reading Scales 305
12.6.2 The Composite Reading Scale 306

12.7 Partitioning of the Estimation Error Variance 307
12.8 Reading Teacher Questionnaires 308

Chapter 13 Data Analysis for the U.S. History Assessment
Bruce A. Kaplan, Eiji Muraki, John J. Ferris, and Steve Wang 309

13.1 Introduction 309
13.2 Description of Student Sample, Items, and Assessment Booklets 310
13.3 Item Analyses 312

13.3.1 Conventional Item and Test Analyses 312
13.3.2 Scoring the Constructed-Response Items 317
13.3.3 Differential Item Functioning 317

13.4 IRT Scaling 319

13.4.1 Item Parameter Estimation 319
13.4.2 Evaluation of Model Fit 320

13.4.3 Linking of Within-Grade Scales 326

13.5 Generation of Plausible Values 326
13.6 The Final U.S. History Scales 329

13.6.1 U.S. History Theme Scales 329
13.6.2 The Composite U.S. History Scale 330

13.7 Partitioning of the Estimation Error Variance 331
13.8 U.S. History Teacher Questionnaires 332

Chapter 14 Data Analysis for the Geography Assessment
Frank Jenkins and Edward Kulick 333

14.1 Introduction 333
14.2 Description of Items and Assessment Booklets 334
14.3 Item Analyses 337

14.3.1 Conventional Item and Test Analyses 337
14.3.2 Scoring the Constructed-Response Items 339
14.3.3 Differential Item Functioning 340

14.4 IRT Scaling 343

14.4.1 Item Parameter Estimation 343
14.4.2 Evaluation of Model Fit 343

14.5 Generation of Plausible Values 351

14.5.1 Principal Components (NSWEEP Program) 351
14.5.2 Conditioning (CGROUP Program) 351
14.5.3 Analysis of Dimensionality 352

14.6 The Final Geography Scales 353

14.6.1 Linking of the Within-Grade Scales Using the TBLT Program 353
14.6.2 The Composite Geography Scales 354

14.7 Partitioning of the Estimation Error Variance 355
14.8 Geography Teacher Questionnaires 356

Chapter 15 Data Analysis for the Long-Term Trend Reading Assessment
Hua-hua Chang, John R. Donoghue, Minhwei Wang,
Lois H. Worthington, and David S. Freund 357

15.1 Introduction 357

15.2 Item Analysis for the Reading Long-Term Trend Assessment ... 361
15.3 Treatment of Constructed-Response Items ... 363
15.4 IRT Scaling for the Reading Long-Term Trend Assessment ... 364

15.4.1 Item Parameter Estimation ... 364
15.4.2 Evaluation of Model Fit ... 365

15.5 Generation of Plausible Values ... 368
15.6 The Final Reading Long-Term Trend Scale ... 369
15.7 Partitioning of the Estimation Error Variance ... 370

Chapter 16 Data Analysis for the Long-Term Trend Mathematics Assessment
Edward Ip, Frank Jenkins, and Edward Kulick ... 373

16.1 Introduction ... 373
16.2 Item Analysis for the Mathematics Long-Term Trend Assessment ... 377
16.3 IRT Scaling for the Mathematics Long-Term Trend Assessment ... 380

16.3.1 Item Parameter Estimation ... 380

16.4 Derived Background Variables
16.5 Generation of Plausible Values ... 384
16.6 The Final Mathematics Long-Term Trend Scale ... 384
16.7 Partitioning of the Estimation Error Variance ... 386

Chapter 17 Data Analysis for the Long-Term Trend Science Assessment
Spencer S. Swinton, Nancy L. Allen, Steven P. Isham, and C. Janet Chen ... 387

17.1 Introduction ... 387
17.2 Item Analysis for the Science Long-Term Trend Assessment ... 392
17.3 IRT Scaling for the Science Long-Term Trend Assessment ... 394

17.3.1 Item Parameter Estimation ... 394
17.3.2 Derived Background Variables ... 398

17.4 Generation of Plausible Values ... 398
17.5 The Final Science Long-Term Trend Scale ... 398
17.6 Partitioning of the Estimation Error Variance ... 400

Chapter 18 Data Analysis for the Long-Term Trend Writing Assessment
James E. Carlson, Bruce A. Kaplan, Eiji Muraki, and Steve Wang ... 401

18.1 Introduction ... 401
18.2 Long-Term Trend Writing Data Analysis ... 403

18.2.1 Primary Trait Scoring of the Writing Tasks and Measures of Scorer Reliability 404

18.3 Item Analysis for the Writing Long-Term Trend Assessment 405
18.4 IRT Scaling for the Writing Long-Term Trend Assessment 405

18.4.1 Item Parameter Estimation Using the Generalized Partial Credit Model 410

18.5 Generation of Plausible Values 411
18.6 The Final Writing Long-Term Trend Scale 412
18.7 Partitioning of Estimation Error Variance 413
18.8 Analysis of Writing Mechanics and Overall Fluency 414

Chapter 19 Statistical Summary of the 1994 NAEP Samples
Bruce A. Kaplan and Phillip T. Leung 417

19.1 Introduction 417
19.2 Measurement Instruments 417
19.3 Sample Characteristics 418
19.4 Population Estimates 419

Appendix A Participants in the Objectives and Item Development Process 497

Appendix B Reporting Subgroups, Composite and Derived Comm Background Variables,
Composite and Derived Reporting Variables 509

Appendix C Conditioning Variables and Contrast Coding 529

Appendix D IRT Parameters 687

Appendix E Estimation Error Variance by Gender and Race/Ethnicity 767

Appendix F Setting the NAEP Achievement Levels for the 1994 Reading Assessment 787

Appendix G Setting the NAEP Achievement Levels for the 1994 U.S. History and Geography Assessments . 819

Appendix H Constructed-Response Item Score Statistics 857

Appendix I Differential Item Functioning (DIF) Results 877

Appendix J Correction of the NAEP Program Documentation Error 883

Appendix K The Information Weighting Error 913

References Cited in Text 947

THE NAEP 1994 TECHNICAL REPORT

◆ List of Tables ◆

Table 1-1 NAEP 1994 Student Samples ... 8
1-2 National Assessment of Educational Progress Subject Areas, Grades, and Ages Assessed: 1969-1994 ... 12
1-3 Main Assessment Reading Booklet Design, Age 13/Grade 8 and Age 17/Grade 12 ... 22
1-4 Main Assessment Reading Booklet Design, Age 9/Grade 4 ... 23

Table 2-1 Percentage Distribution of Items by Reading Purpose as Specified in the NAEP Reading Framework ... 39
2-2 Percentage Distribution of Items by Reading Stance as Specified in the NAEP Reading Framework ... 39
2-3 Distribution of Items for the 1994 Reading Assessment ... 42
2-4 Percent Distribution of Assessment Time by Grade and Reading Purpose ... 42
2-5 Percentage Distribution of Assessment Time by Grade and Reading Stance* ... 43
2-6 Percentage Distribution of Assessment Time Across U.S. History Themes as Specified in the History Framework ... 47
2-7 Distribution of Assessment Time Across U.S. History Periods as Specified in the Assessment Framework ... 48
2-8 Distribution of Assessment Time Across Cognitive Dimensions as Specified in the U.S. History Assessment Framework ... 48
2-9 "Average" Number of Items in 25-Minute U.S. History Blocks ... 50
2-10 Distribution of Questions for the 1994 U.S. History Assessment ... 51
2-11 Distribution of Assessment Time by U.S. History Item Format ... 51
2-12 Distribution of Assessment Time by U.S. History Theme ... 52
2-13 Distribution of Assessment Time by U.S. History Period ... 52
2-14 Distribution of Assessment Time by U.S. History Cognitive Dimension ... 53
2-15 Percentage of Assessment Time Across Content Areas as Specified in the Geography Framework ... 58
2-16 Percentage of Assessment Time Across Cognitive Dimensions as Specified in the Geography Framework ... 59
2-17 "Average" Number of Items in 25-Minute Geography Blocks ... 60
2-18 Distribution of Items for the 1994 Geography Assessment ... 61
2-19 Distribution of Assessment Time by Geography Item Format ... 62
2-20 Distribution of Assessment Time by Geography Content Area ... 62
2-21 Distribution of Assessment Time and Item Pool by Geography Cognitive Dimension ... 63

Table 3-1 Assessment Type by Age Class and Assessment Period ... 70
3-2 Geographic Regions Used for Stratification ... 71
3-3 The Sampling Strata and the Number of Noncertainty Strata in Each ... 72
3-4 Grade Definition of School Eligibility for Frame Inclusion and Frame Sizes, Main and Long-Term Trend Samples ... 74
3-5 School Sample Sizes, Refusals, and Substitutes for the Main Samples ... 77
3-6 School Sample Sizes, Refusals, and Substitutes for the Long-Term Trend Samples ... 78
3-7 Number of Students Per School for Each Assessment ... 81

3-8 Student Exclusion Rates by Age Class and School Type, Unweighted 82
3-9 Student Participation Rates by Age Class and School Type, Unweighted 83
3-10 Overall Participation Rates (School and Student Combined) by Age Class, Unweighted 84
3-11 Weighted Participation Rates by Age Class and Subject, Main Samples 85
3-12 Weighted Participation Rates by Age Class and Subject, Long-Term Trend Samples 85

Table 4-1 Main Sample Booklet Contents and Number of Booklets Administered Age 9/Grade 4, Reading 89
4-2 Cognitive and Noncognitive Block Information, Age 9/Grade 4, Reading 90
4-3 Main Sample Booklet Contents and Number of Booklets Administered Age 13/Grade 8, Reading 91
4-4 Cognitive and Noncognitive Block Information, Age 13/Grade 8, Reading 92
4-5 Main Sample Booklet Contents and Number of Booklets Administered Age 17/Grade 12, Reading 93
4-6 Cognitive and Noncognitive Block Information, Age 17/Grade 12, Reading 94
4-7 Main Sample Booklet Contents and Number of Booklets Administered Age 9/Grade 4, U.S. History 96
4-8 Cognitive and Noncognitive Block Information, Age 9/Grade 4, U.S. History 97
4-9 Main Sample Booklet Contents and Number of Booklets Administered Age 13/Grade 8 and Age 17/Grade 12, U.S. History 98
4-10 Cognitive and Noncognitive Block Information, Age 13/Grade 8, U.S. History 99
4-11 Cognitive and Noncognitive Block Information, Age 17/Grade 12, U.S. History 100
4-12 Main Sample Booklet Contents and Number of Booklets Administered Age 9/Grade 4, Geography 102
4-13 Cognitive and Noncognitive Block Information, Age 9/Grade 4, Geography 103
4-14 Main Sample Booklet Contents and Number of Booklets Administered Age 13/Grade 8 and Age 17/Grade 12, Geography 104
4-15 Cognitive and Noncognitive Block Information, Age 13/Grade 8, Geography 105
4-16 Cognitive and Noncognitive Block Information, Age 17/Grade 12, Geography 106
4-17 Long-Term Trend Sample Booklet Contents and Number of Booklets Administered Age Class 9 107
4-18 Long-Term Trend Sample Booklet Contents and Number of Booklets Administered Age Class 13 108
4-19 Long-Term Trend Sample Booklet Contents and Number of Booklets Administered Age Class 17 108
4-20 Long-Term Trend Sample Block Information. Age Class 9 109
4-21 Long-Term Trend Sample Block Information, Age Class 13 110
4-22 Long-Term Trend Sample Block Information, Age Class 17 111

Table 5-1 Comparison of Student and School Response Rates (Unweighted) for Main Assessments, 1990-1994 123
5-2 Response Data for Main Assessment Questionnaires 124
5-3 Comparison of Student and School Response Rates (Unweighted) for Long-Term Trend Assessments, 1990-1994 130

Table 6-1 Processing Totals for the 1994 NAEP Assessment 135
6-2 1994 NAEP Assessment Schedule of Printing and Processing Activities 135

Table 7-1 Number of Constructed-Response Items, Long-Term Trend Assessments 156
7-2 Number of Constructed-Response Items, 1993-94 Main Assessments 156
7-3 1993-94 NAEP Assessment Professional Scoring - Readers and Dates 158
7-4 1993-94 NAEP Long-Term Trend Interreader Reliability Ranges 160
7-5 1993-94 NAEP Assessment, Professional Scoring - Holistic Writing Scoring Reliability Figures 162
7-6 1993-94 NAEP U.S. History Assessment, Interreader Reliability Ranges 166
7-7 1993-94 NAEP Geography Assessment Interreader Reliability Ranges 169
7-8 1993-94 NAEP Reading Assessment Interreader Reliability Ranges 172
7-9 1993-94 NAEP Assessment, 1992 National Reading Booklets and Blocks Rescored on Paper-Based and Image Systems 173

Table 8-1 Summary of Quality Control Error Analysis for NAEP 1994 Data Entry 181

Table 10-1 Distribution of Populations of Eligible Students Based on Full Weighted Sample of Eligible Schools, 1994 Main NAEP Samples 216
10-2 Distribution of Populations of Eligible Students Based on Weighted Sample of Participating Schools with School Nonresponse Adjustments, 1994 Main NAEP Samples . . 217
10-3 Distribution of Populations of Eligible Students Based on Assessed and Absent Students from Participating Schools, 1994 Main NAEP Samples 218
10-4 Distribution of Populations of Eligible Students Based on Assessed Students from Participating Schools with Student Nonresponse Adjustments Applied 1994 Main NAEP Samples 218
10-5 Distribution of Populations of Eligible Students in Participating Schools Based on Invited Students from Participating Schools Using Base Weights Unadjusted for School Nonresponse 1994 Main NAEP Reading Sample 220
10-6 Distribution of Populations of Eligible Students Based on Invited Students from Participating Schools Using Weights Adjusted for School Nonresponse But Not Student Nonresponse 1994 Main NAEP Reading Sample 221
10-7 Distribution of Populations of Eligible Students Based on Assessed Students from Participating Schools with Student Nonresponse Adjustments Applied 1994 Main NAEP Reading Sample 222
10-8 Weighted Distribution of Absent Students by Nature of Absenteeism for All Age Classes 1994 Reading Assessment 223
10-9 Distribution of Populations of Eligible Students Based on Trimmed Weights of Assessed Students in Participating Schools, 1994 Main NAEP Reading Sample 225
10-10 Major Subgroups for Poststratification in 1994 226
10-11 Distribution of Final Student Weights, Long-Term Trend Samples 228
10-12 Distribution of Final Student Weights, Main Samples 229
10-13 Design Effects by Demographic Subgroup and Grade for Mean U.S. History Proficiencies 239
10-14 Design Effects by Demographic Subgroup and Grade for Mean Geography Proficiencies 240
10-15 Design Effects by Demographic Subgroup and Grade for Mean Reading Proficiencies 241
10-16 Design Effects by Demographic Subgroup and Age Averaged Over Mean Reading and Writing Proficiencies for the Reading and Writing Long-Term Trend Samples 242
10-17 Design Effects by Demographic Subgroup and Age Averaged Over Mean Mathematics and Science Proficiencies for the Mathematics and Science Long-Term Trend Samples 243

10-18 Average Design Effects by Demographic Subgroup for Mean State Reading Proficiencies Averaged Across State Samples 244
10-19 Within-Grade Mean, Median, and Upper Quartile of the Distribution of Design Effects for National Main Assessments by Subject Area and Across Subject Areas 245
10-20 Mean, Median, and Upper Quartile of the Across-State Average Design Effects for Mean State Reading Proficiency (Distribution Across Demographic Subgroups) 246
10-21 Mean, Median, and Upper Quartile of the Distribution of Design Effects for the Long-Term Trend Assessment Samples 246

Table 11-1 Estimation Error Variance and Related Coefficients for the 1992 Grade 4 Reading Composite (Based on Five Plausible Values) 263

Table 12-1 NAEP 1994 Reading Student Samples 269
12-2 1994 Reading Cross-Sectional Blocks and Items Common to the 1992 Assessment 270
12-3 Item Type Composition of Blocks Used in the Reading Main Assessment 273
12-4 Numbers of Scaled Reading Items Common Across Grade Levels 274
12-5 Descriptive Statistics for Item Blocks by Position Within Booklet and Over All Occurrences for the Reading Main Sample, Grade 4 275
12-6 Descriptive Statistics for Item Blocks by Position Within Booklet and Over All Occurrences for the Reading Main Sample, Grade 8 276
12-7 Descriptive Statistics for Item Blocks by Position Within Booklet and Over All Occurrences for the Reading Main Sample, Grade 12 277
12-8 DIF Category by Grade for Dichotomous Items 280
12-9 DIF Category by Grade for Polytomous Items 281
12-10 Reanalysis of 1992 Reading Assessment: Coefficients of Linear Transformations of the Scales from the Calibrating Scale Units to the Units of Reporting Proficiency 283
12-11 Items Deleted from the Final Scaling 289
12-12 Recoding of Polytomous Items for Scaling 292
12-13 Age 9/Grade 4 Items Scaled Separately by Assessment Year 300
12-14 Age 13/Grade 8 Items Scaled Separately by Assessment Year 301
12-15 Age 17/Grade 12 Items Scaled Separately by Assessment Year 302
12-16 Proportion of Proficiency Variance Accounted for by the Conditioning Model for the Reading Main Assessment 304
12-17 Conditional Correlations and Variances from Conditioning (CGROUP) 304
12-18 Marginal Correlations of Reading Scales 305
12-19 Coefficients of Linear Transformations of the Purpose-for-Reading Scales from the Calibrating Scale Units to the Units of Reporting Proficiency 306
12-20 Weights Used for Each Purpose-for-Reading Scale to Form the Reading Composite 306
12-21 Means and Standard Deviations on the Reading Composite Scale 307
12-22 Estimation Error Variance and Related Coefficients for the Reading Main Assessment 308

Table 13-1 NAEP 1994 U.S. History Student Samples 310
13-2 Item Type Composition of Blocks Used in the U.S. History Assessment 313
13-3 Numbers of Scaled U.S. History Items by Grade Levels 314
13-4 Descriptive Statistics for Item Blocks for the U.S. History Sample, Grade 4 314
13-5 Descriptive Statistics for Item Blocks for the U.S. History Sample, Grade 8 315
13-6 Descriptive Statistics for Item Blocks for the U.S. History Sample, Grade 12 316
13-7 U.S. History DIF Categories by Grade for Dichotomous Items 318

13-8 U.S. History DIF Categories by Polytomous Items. 319
13-9 Items from the 1994 U.S. History Main Analysis Receiving Special Treatment 321
13-10 Linear Transformation Constants for the U.S. History Between Grade Linking 326
13-11 Proportion of Proficiency Variance Accounted for by the Conditioning Model for the U.S. History Assessment 327
13-12 Conditional Correlations and Residual Variances from Conditioning (CGROUP) 328
13-13 Marginal Correlations of U.S. History Scales 329
13-14 Coefficients of Linear Transformations of the U.S. History Themes Scales from the Calibrating Scale Units to the Units of the U.S. History Scale 330
13-15 Weights Used for Each U.S. History Theme Scale to Form the U.S. History Composite 331
13-16 Means and Standard Deviations on the U.S. History Composite Scale 331
13-17 Estimation Error Variance and Related Coefficients for the U.S. History Assessment 332

Table 14-1 NAEP 1994 Geography Student Samples 334
14-2 Numbers of Scaled Geography Items Common Across Grade Levels, by Content Area Scale 335
14-3 Item Type Composition of Blocks Used in the Geography Assessment 336
14-4 Numbers of Items in Scales 337
14-5 Descriptive Statistics for Item Blocks from the Geography Main Sample, Grade 4 338
14-6 Descriptive Statistics for Item Blocks from the Geography Main Sample, Grade 8 339
14-7 Descriptive Statistics for Item Blocks from the Geography Main Sample, Grade 12 339
14-8 DIF Category by Grade for Dichotomous Items 341
14-9 DIF Category by Grade for Polytomous Items 342
14-10 Items from the 1994 Geography Main Analysis Receiving Special Treatment 345
14-11 Proportion of Proficiency Variance Accounted for by the Conditioning Model for the Geography Main Assessment 351
14-12 Conditional Correlations of Geography from Conditioning (CGROUP) 352
14-13 Marginal Correlations of Geography Scales 353
14-14 Coefficients of Linear Transformations of the Geography Scales from the Calibrating Scale Units to the Units of Reporting Proficiency 354
14-15 Weights Used for Each Scale to Form the Geography Composite 354
14-16 Means and Standard Deviations on the Geography Composite Scale 355
14-17 Estimation Error Variance and Related Coefficients for the Geography Main Assessment 355

Table 15-1 NAEP 1994 Long-Term Trend Reading Student Samples 358
15-2 NAEP Reading Samples Contributing to 1994 Long-Term Trend Results, 1971-1994 359
15-3 Numbers of Scaled Reading Long-Term Trend Items Common Across Ages 360
15-4 Numbers of Scaled Reading Long-Term Trend Items Common Across Assessments 361
15-5 Descriptive Statistics for Item Blocks in the Reading Long-Term Trend Samples 362
15-6 Items Deleted from the Reading Long-Term Trend Analysis 364
15-7 Items Calibrated Separately by Assessment Year in the Reading Long-Term Trend Analysis 368
15-8 Proportion of Proficiency Variance Accounted for by the Conditioning Model for the Reading Long-Term Trend Assessment 369
15-9 Means and Standard Deviations on the Reading Long-Term Trend Scale 370
15-10 Estimation Error Variance and Related Coefficients for the Reading Long-Term Trend Assessment 371

Table 16-1 NAEP 1994 Mathematics Student Samples 374

16-2 NAEP Mathematics Samples Contributing to 1994 Long-Term Trend Results, 1973-1994 375
16-3 Numbers of Scaled Mathematics Long-Term Trend Items Common Across Ages 376
16-4 Numbers of Scaled Mathematics Long-Term Trend Items Common Across Assessments 376
16-5 Descriptive Statistics for Item Blocks in the Mathematics Long-Term Trend Samples (1994) .. 378
16-6 Descriptive Statistics for Item Blocks in the Mathematics Long-Term Trend Samples (1992) .. 379
16-7 Items Deleted from the Age 9 Mathematics Long-Term Trend Analysis 381
16-8 Items Deleted from the Age 13 Mathematics Long-Term Trend Analysis 382
16-9 Items Deleted from the Age 17 Mathematics Long-Term Trend Analysis 383
16-10 Proportion of Proficiency Variance Accounted for by the Conditioning Model for the Mathematics Long-Term Trend Assessment 384
16-11 Means and Standard Deviations on the Mathematics Long-Term Trend Proficiency Scale 385
16-12 Estimation Error Variance and Related Coefficients for the Mathematics Long-Term Trend Assessment 386

Table 17-1 NAEP 1994 Science Student Samples 388
17-2 NAEP Science Samples Contributing to 1994 Long-Term Trend Results, 1970-1994 390
17-3 Numbers of Scaled Science Long-Term Trend Items Common Across Ages 391
17-4 Numbers of Scaled Science Long-Term Trend Items Common Across Assessments 392
17-5 Descriptive Statistics for Item Blocks in the Science Long-Term Trend Samples (1994) 393
17-6 Items Deleted from the Paced-Tape Science Long-Term Trend Analysis 395
17-7 Proportion of Proficiency Variance Accounted for by the Conditioning Model for the Science Long-Term Trend Assessment 398
17-8 Means and Standard Deviations on the Science Long-Term Trend Scale 399
17-9 Estimation Error Variance and Related Coefficients for the Science Long-Term Trend Assessment 400

Table 18-1 NAEP Writing Samples Contributing to 1994 Long-Term Trend Results, 1984-1994 402
18-2 NAEP Writing Student Samples 403
18-3 Assignment of 1984-94 Writing Long-Term Trend Items in 1984, 1988, 1990, 1992, and 1994 404
18-4 Percentages of Exact Score Agreement and Interrater Reliability for the Primary Trait Scoring of the Writing Long-Term Trend Assessment Items 406
18-5 Descriptive Statistics for Writing Prompts, 1994 Writing Long-Term Trend Samples 408
18-6 Age/Grade Sample Sizes for Scaling Items Used in Long-Term Trends in Writing Performance 409
18-7 Nonoverlapping Sets of Items for the Writing Long-Term Trend Assessment 411
18-8 Proportion of Proficiency Variance Accounted for by the Conditioning Model for the Writing Long-Term Trend Assessment 412
18-9 Means and Standard Deviations on the Writing Long-Term Trend Scale 413
18-10 Estimation Error Variance and Related Coefficients for the Writing Long-Term Trend Assessment 413
18-11 Sample Sizes for Mechanics Scoring 414
18-12 Sample Sizes for Holistic Scoring 415

Table 19-1 Measurement Instruments Developed for 1994 NAEP 422
19-2 Number of Items Administered, by Sample and Age Class 423
19-3 Number of Distinct Cognitive Items, by Sample and Age Class 425
19-4 School Characteristics in Main Samples 426
19-5 School Characteristics in Long-Term Trend Samples 428

19-6 Numbers of Responses to Teacher Questionnaires and Students Matched with Teacher Data 430
19-7 Numbers of Assessment Sessions by Sample, Type of Session, and Age Class 431
19-8 Number of Students Assessed and Excluded by Sample and Age Class 432
19-9 Number of Students in the Reading Main Sample by Type of Eligibility and Subgroup Classification, Age 9/Grade 4 433
19-10 Number of Students in the U.S. History Main Sample by Type of Eligibility and Subgroup Classification, Age 9/Grade 4 434
19-11 Number of Students in the Geography Main Sample by Type of Eligibility and Subgroup Classification, Age 9/Grade 4 435
19-12 Number of Students in the Reading Main Sample by Type of Eligibility and Subgroup Classification, Age 13/Grade 8 436
19-13 Number of Students in the U.S. History Main Sample by Type of Eligibility and Subgroup Classification, Age 13/Grade 8 437
19-14 Number of Students in the Geography Main Sample by Type of Eligibility and Subgroup Classification, Age 13/Grade 8 438
19-15 Number of Students in the Reading Main Sample by Type of Eligibility and Subgroup Classification, Age 17/Grade 12 439
19-16 Number of Students in the U.S. History Main Sample by Type of Eligibility and Subgroup Classification, Age 17/Grade 12 440
19-17 Number of Students in the Geography Main Sample by Type of Eligibility and Subgroup Classification, Age 17/Grade 12 441
19-18 Number of Students in the Reading and Writing Long-Term Trend Sample by Type of Eligibility and Subgroup Classification, Age 9/Grade 4 442
19-19 Number of Students in the Reading and Writing Long-Term Trend Sample by Type of Eligibility and Subgroup Classification, Age 13/Grade 8 443
19-20 Number of Students in the Reading and Writing Long-Term Trend Sample by Type of Eligibility and Subgroup Classification, Age 17/Grade 11 444
19-21 Number of Students in the Mathematics and Science Long-Term Trend Sample by Type of Eligibility and Subgroup Classification, Age 9/Grade 4 445
19-22 Number of Students in the Mathematics and Science Long-Term Trend Sample by Type of Eligibility and Subgroup Classification, Age 13/Grade 8 446
19-23 Number of Students in the Mathematics and Science Long-Term Trend Sample by Type of Eligibility and Subgroup Classification, Age 17/Grade 11 447
19-24 Number of Excluded Students in Main Samples by Type of Eligibility and Subgroup Classification, Age 9/Grade 4 448
19-25 Number of Excluded Students in Main Samples by Type of Eligibility and Subgroup Classification, Age 13/Grade 9 449
19-26 Number of Excluded Students in Main Samples by Type of Eligibility and Subgroup Classification, Age 17/Grade 12 450
19-27 Number of Excluded Students in Long-Term Trend Samples by Type of Eligibility and Subgroup Classification, Age 9/Grade 4 451
19-28 Number of Excluded Students in Long-Term Trend Samples by Type of Eligibility and Subgroup Classification, Age 13/Grade 8 452
19-29 Number of Excluded Students in Long-Term Trend Samples by Type of Eligibility and Subgroup Classification, Age 17/Grade 12 453
19-30 Number of Students by Sample and Age Class 454
19-31 Number of Students Assessed in the Reading Main Sample, Age 9/Grade 4 455

19-32 Number of Students Assessed in the U.S. History Main Sample, Age 9/Grade 4 456
19-33 Number of Students Assessed in the Geography Main Sample, Age 9/Grade 4 457
19-34 Number of Students Assessed in the Reading Main Sample, Age 13/Grade 8 458
19-35 Number of Students Assessed in the U.S. History Main Sample, Age 13/Grade 8 459
19-36 Number of Students Assessed in the Geography Main Sample, Age 13/Grade 8 460
19-37 Number of Students Assessed in the Reading Main Sample, Age 17/Grade 12 461
19-38 Number of Students Assessed in the U.S. History Main SampleAge 17/Grade 12 462
19-39 Number of Students Assessed in the Geography Main Sample, Age 17/Grade 12 463
19-40 Number of Students Assessed in the Reading and Writing Long-Term Trend Sample Age 9/Grade 4 464
19-41 Number of Students Assessed in the Reading and Writing Long-Term Trend Sample Age 13/Grade 8 465
19-42 Number of Students Assessed in the Reading and Writing Long-Term Trend Sample Age 17/Grade 11 466
19-43 Number of Students Assessed in the Mathematics and Science Long-Term Trend Sample 467
19-44 Weighted Percentage of Students in the Reading Main Sample by Type of Eligibility and Subgroup Classification, Age 9/Grade 4 468
19-45 Weighted Percentage of Students in the U.S. History Main Sample by Type of Eligibility and Subgroup Classification, Age 9/Grade 4 469
19-46 Weighted Percentage of Students in the Geography Main Sample by Type of Eligibility and Subgroup Classification, Age 9/Grade 4 470
19-47 Weighted Percentage of Students in the Reading Main Sample by Type of Eligibility and Subgroup Classification, Age 13/Grade 8 471
19-48 Weighted Percentage of Students in the U.S. History Main Sample by Type of Eligibility and Subgroup Classification, Age 13/Grade 8 472
19-49 Weighted Percentage of Students in the Geography Main Sample by Type of Eligibility and Subgroup Classification, Age 13/Grade 8 473
19-50 Weighted Percentage of Students in the Reading Main Sample by Type of Eligibility and Subgroup Classification, Age 17/Grade 12 474
19-51 Weighted Percentage of Students in the U.S. History Main Sample by Type of Eligibility and Subgroup Classification, Age 17/Grade 12 475
19-52 Weighted Percentage of Students in the Geography Main Sample by Type of Eligibility and Subgroup Classification, Age 17/Grade 12 476
19-53 Weighted Percentage of Students in the Reading and Writing Long-Term Trend Sample by Type of Eligibility and Subgroup Classification, Age 9/Grade 4 477
19-54 Weighted Percentage of Students in the Reading and Writing Long-Term Trend Sample by Type of Eligibility and Subgroup Classification, Age 13/Grade 8 478
19-55 Weighted Percentage of Students in the Reading and Writing Long-Term Trend Sample by Type of Eligibility and Subgroup Classification, Age 17/Grade 11 479
19-56 Weighted Percentage of Students in the Mathematics and Science Long-Term Trend Sample by Type of Eligibility and Subgroup Classification, Age 9/Grade 4 480
19-57 Weighted Percentage of Students in the Mathematics and Science Long-Term Trend Sample by Type of Eligibility and Subgroup Classification, Age 13/Grade 8 481
19-58 Weighted Percentage of Students in the Mathematics and Science Long-Term Trend Sample by Type of Eligibility and Subgroup Classification, Age 17/Grade 11 482
19-59 Weighted Percentage of Excluded Students in the Main Samples by Type of Eligibility and Subgroup Classification, Age 9/Grade 4 483

19-60 Weighted Percentage of Excluded Students in the Main Samples by Type of Eligibility and Subgroup Classification, Age 13/Grade 8 484
19-61 Weighted Percentage of Excluded Students in the Main Samples by Type of Eligibility and Subgroup Classification, Age 17/Grade 12 485
19-62 Weighted Percentage of Excluded Students in the Long-Term Trend Samples by Type of Eligibility and Subgroup Classification, Age 9/Grade 4 486
19-63 Weighted Percentage of Excluded Students in the Long-Term Trend Samples by Type of Eligibility and Subgroup Classification, Age 13/Grade 8 487
19-64 Weighted Percentage of Excluded Students in the Long-Term Trend Samples by Type of Eligibility and Subgroup Classification, Age 17/Grade 11 488
19-65 Number of Students in the Main Samples, with Proficiency Scores, by Type of Eligibility 489
19-66 Number of Students in the Long-Term Trend Samples 490
19-67 Raw Counts and Total Unweighted Numbers of IEP and LEP Students, Age 9/Grade 4 491
19-68 Raw Counts and Total Unweighted Numbers of IEP and LEP Students, Age 13/Grade 8 492
19-69 Raw Counts and Total Unweighted Numbers of IEP and LEP Students, Age 17/Grade 12 493
19-70 Weighted Percentages and Total Weighted Numbers of IEP and LEP Students Age 9/Grade 4 494
19-71 Weighted Percentages and Total Weighted Numbers of IEP and LEP Students Age 13/Grade 8 495
19-72 Weighted Percentages and Total Weighted Numbers of IEP and LEP Students Age 17/Grade 12 496

THE NAEP 1994 TECHNICAL REPORT

◆ List of Figures ◆

Figure 2-1 Description of Reading Stances . . . 37
2-2 Description of Purposes for Reading . . . 38

Figure 5-1 Instructions for Excluding Students (Main Assessments) . . . 121
5-2 Instructions for Excluding Students (Long-Term Trend Assessments) . . . 129

Figure 11-1 Example Cross-Sectional Dichotomous Item (R016102, Age 13/Grade 8) Exhibiting Good Model Fit . . . 254
11-2 Example Cross-Sectional Polytomous Item (R015804, Age 9/Grade 4) Exhibiting Good Model Fit . . . 255

Figure 12-1 Example Cross-Sectional Dichotomous Item (R016102, Age 13/Grade 8) Exhibiting Good Model Fit . . . 286
12-2 Example Cross-Sectional Polytomous Item (R015804, Age 9/Grade 4) Exhibiting Good Model Fit . . . 287
12-3 Example Cross-Sectional Polytomous Item (R016603, Age 17/Grade 12) Exhibiting Unacceptably Poor Model Fit . . . 288
12-4 Example Cross-Sectional Polytomous Item (R015707, Age 9/Grade 4) Exhibiting Poor Model Fit . . . 290
12-5 Example Cross-Sectional Polytomous Item (R015707, Age 9/Grade 4) After Collapsing Categories 2 and 3 . . . 291
12-6 Example Short-Term Trend Dichotomous Item (R013102, Age 13/Grade 8) Demonstrating Differential Item Functioning Across Assessment Years 1992 and 1994 . . . 297
12-7 Example Short-Term Trend Dichotomous Item (R013102, Age 13/Grade 8) Fitting Separate Item Response Functions for Each Assessment Year . . . 298
12-8 Example Short-Term Trend Dichotomous Item (R013405, Age 17/Grade 12) Demonstrating Acceptably Small Differential Item Functioning Across Assessment Years 1992 and 1994 . . . 299
Figure 13-1 An Example of a Multiple-Choice (Dichotomous) Item (H053402, Grade 12), Exhibiting Poor Model Fit . . . 322
13-2 An Example of a Multiple-Choice (Dichotomous) Item (H037301, Grade 8) Exhibiting Good Model Fit . . . 323
13-3 An Example of a Polytomous Item (H028201, Grade 4) Exhibiting Good Model Fit . . . 324
13-4 An Example of a Polytomous Item (H043501, Grade 8) Exhibiting Poor Model Fit . . . 325

Figure 14-1 Example Dichotomous Item (G013701) Exhibiting Good Model Fit . . . 346
14-2 Example Dichotomous Item (G017201) Exhibiting Poor Model Fit . . . 347
14-3 Example Polytomous Item (G012503) Exhibiting Poor Model Fit . . . 348
14-4 Example Dichotomous Item (G012503) Exhibiting Good Model Fit After Polytomous Categories (Seen in 14-3) Were Collapsed . . . 349
14-5 Example Polytomous Item (G018301) Exhibiting Good Model Fit . . . 350

Figure 15-1 Example Long-Term Trend Item (N003704, Age 17) Demonstrating Differential Item Functioning Across Assessment Years 1992 and 1994 366
15-2 Example Long-Term Trend Item (N003704, Age 17) Fitting Separate Item Response Functions for Each Assessment Year 367

Figure 17-1 Example Long-Term Trend Item (N4063032, Age 13) Demonstrating Differential Item Functioning Across Assessment Years 1990 and 1992 396
17-2 Example Long-Term Trend Item (N4063032, Age 13) Exhibiting Good Model Fit Across Assessment Years 1992 and 1994 397

◆ Acknowledgments ◆

The design, development, administration, analysis, and reporting of the 1994 National Assessment of Educational Progress (NAEP) was a collaborative effort among staff from the National Center for Education Statistics (NCES), the National Assessment Governing Board (NAGB), the Council of Chief State School Officers (CCSSO), Educational Testing Service (ETS), Westat, and National Computer Systems (NCS). This report documents the NAEP design, administration, and data analysis procedures and provides insight about the rationale behind the technical decisions made. The development of this report and of the national assessment program is the result of the considerable knowledge, experience, creativity, and dedication of many individuals. It is fitting that these individuals are acknowledged for their contribution to NAEP.

The 1994 NAEP was funded through the National Center for Education Statistics in the Office of Educational Research and Improvements of the U.S. Department of Education. Emerson Elliott, former NCES commissioner, provided consistent support and guidance. The staff worked closely and collegially with ETS, Westat, and NCS staff and played a crucial role in all aspects of the program. We would like to especially recognize and thank Shi-Chang Wu for his significant assistance and support in coordinating the organization and technical review of this report. We are also particularly grateful for the contributions of Gary Phillips, Peggy Carr, Mary Frase, Arnie Goldstein, Stephen Gorman, Andrew Kolstad, Larry Ogle, Mary Rollefson, Sharif Shikrani, Maureen Treacy, and Sheida White.

We would also like to gratefully acknowledge Susan Ahmed, Ralph Lee, Andrew Kolstad, Stephen Gorman, and Shi-Chang Wu of NCES, Mary Lyn Bourque of NAGB, Jamal Abedi of UCLA, and Don McLaughlin of American Institutes for Research (AIR), for conducting a technical review of this report.

Special thanks also go to the members of the National Assessment Governing Board (NAGB) and the NAGB staff who provided advice and guidance in the preparation of this report, particularly Mary Lyn Bourque, who provided information for some of the report appendices.

ETS management and the Center for the Assessment of Educational Progress (CAEP) have been very supportive of NAEP's technical work. Special thanks go to Nancy Cole as well as to Henry Braun and Charles Davis of ETS research management, and Archie Lapointe, Jules Goodison, Paul Williams, Steve Lazer, John Mazzeo, Clyde Reese, and Kim Whittington of CAEP. Significant contributions to the project were also received from Nada Ballator, Alexandra Beatty, Jay Campbell, Patricia Donahue, Claudia Gentile, Janet Johnson, Andrew Latham, Mary Michaels, Christine O'Sullivan, and Hilary Persky.

The guidance of the NAEP Design and Analysis Committee on technical aspects of NAEP has been outstanding. During the period of analysis of the 1994 data, the members were Sylvia Johnson (chair), Albert Beaton, Jeri Benson, William Cooley, Jeremy Finn, Paul Holland, Huynh Huynh, Gaea Leinhardt, David Lohman, Bengt Muthén, Anthony Nitko, Ingram Olkin, Tej Pandey, Juliet Shaffer, and Hariharan Swaminathan. We were saddened by the untimely death of Clifford Clogg, whose service to NAEP will not be forgotten.

The design and data analysis of the 1994 National Assessment was primarily the responsibility of the NAEP research and data analysis staff at ETS with significant contributions from NAEP management, Westat, and NCS staffs. In addition to managing day-to-day data analytic operations, these staff members have made many innovative statistical and psychometric contributions. Major contributions were made by Eugene Johnson, James Carlson, Hua-hua Chang, John Donoghue, Edward Ip, Frank Jenkins, John Mazzeo, Eiji

Muraki, and Spencer Swinton. Jo-lin Liang was exceptional in her role as technical assistant. Eugene Johnson, Robert Mislevy, Juliet Shaffer, Neal Thomas, and Ming-mei Wang provided valuable statistical and psychometric advice.

The Divison of Data Analysis and Technology Research at ETS, under the leadership of John Barone, was responsible for developing the operating systems and carrying out the data analyses. David Freund and Alfred Rogers developed and maintained the large and complex NAEP data management systems; Katharine Pashley managed database activities and Patricia O'Reilly was responsible for the creation of the secondary-use data files. Alfred Rogers developed the production versions of key analysis and scaling systems. Special recognition goes to Phillip Leung and Patricia O'Reilly for their expertise in the design and development of the NAEP Web site. Thanks also go to David Freund, Steven Isham, Bruce Kaplan, Debbie Kline, and Edward Kulick, for their continuing roles as leaders and developers of innovative solutions to NAEP data analysis challenges, and to Thomas Jirele, Michael Narcowich, Stephen Szyszkiewicz, and Probal Tahbildar for their contributions to the development of NAEP data analysis software Many members of this division made substantial and important contributions of their time and talent to NAEP data analyses, including C. Janet Chen, John Ferris, Laura Jerry, Phillip Leung, Laura Jenkins, Michael Narcowich, Jennifer Nelson, Norma Norris, Inge Novatkoski, Patricia O'Reilly, Katharine Pashley, Craig Pizzuti, Minhwei Wang, Steve Wang, Lois Worthington, and Hong Zhou.

The staff at Westat contributed their talents and efforts in all areas of the sample design and data collection. Particular recognition is due to Nancy Caldwell and Dianne Walsh for supervising the field operations, and to Keith Rust for developing and supervising the sample design and selection. Thanks are also due to Bridgett Bell, Frances Cohen, Ismael Flores-Cervantes, Brice Hart, Sandra Rieder, Ellen Tubbs, Mark Waksberg, Debra Vivari, Ngoan Vo, Leslie Wallace, Dianne Walsh, and Marianne Whitlock. The exceptional data collection effort of the Westat field staff is also gratefully acknowledged.

Critical to the program was the contribution of National Computer Systems, which has been responsible for the printing, distribution, and processing of the assessment materials. The leadership roles of Judith Moyer and Bradley Thayer are especially acknowledged. Thanks also go to Patrick Bourgeacq, Charles Brungardt, Mary Lou Close, Linda Reynolds, Mathilde Kennel, Paul McRoberts, Tim Robinson, Connie Smith, Patricia Garcia Stearns, and Brent Studer.

Sharon Davis-Johnson also deserves special thanks for coordinating the design of the report's cover.

Special recognition and appreciation go to Debra Kline and Christine Zelenak, editors of this report, for organizing, scheduling, editing, motivating, and ensuring the cohesiveness and correctness of the final report.

Finally, NAEP is grateful to the students and school staff members who participated in the assessment. Without their efforts, there would be no assessment.

Nancy L. Allen
Director of Data Analysis and Scaling
NAEP Research, ETS
December 1996

THE NAEP 1994 TECHNICAL REPORT

Introduction[1]

Juliet Shaffer and James E. Carlson
Educational Testing Service

The 1994 National Assessment of Educational Progress (NAEP) monitored the performance of students in American schools in the subject areas of reading, U.S. history, geography, mathematics, science, and writing. The sample involved more than 135,000 public- and nonpublic-school students who were 9-, 13-, or 17 years old or in grades 4, 8, or 12.

The purpose of this technical report is to provide details on the instrument development, sample design, data collection, and data analysis procedures of the 1994 assessment. Detailed substantive results are not presented here but can be found in a series of NAEP reports on the status of and trends in student performance; several additional reports provide information on how the assessment was designed and implemented. The reader is directed to the following reports for 1994 results and supporting documentation:

- *1994 NAEP Reading: A First Look—Findings from the National Assessment of Educational Progress* (Williams, Reese, Campbell, Mazzeo, & Phillips, 1995)
- *NAEP 1994 U.S. History: A First Look—Findings from the National Assessment of Educational Progress* (Williams, Lazer, Reese, & Carr, 1995)
- *NAEP 1994 Geography: A First Look—Findings from the National Assessment of Educational Progress* (Williams, Reese, Lazer, & Shakrani, 1995)
- *NAEP 1994 Reading Report Card for the Nation and the States: Findings from the National Assessment of Educational Progress and the Trial State Assessments* (Campbell, Donahue, Reese, & Phillips, 1996)
- *NAEP 1994 U.S. History Report Card: Findings from the National Assessment of Educational Progress* (Beatty, Reese, Persky, & Carr, 1996)
- *NAEP 1994 Geography Report Card: Findings from the National Assessment of Educational Progress* (Persky, Reese, O'Sullivan, Lazer, Moore, & Shakrani, 1996).
- *Reading Framework for the 1992 and 1994 National Assessment of Educational Progress* (NAGB, 1994).

---

[1] Juliet Shaffer is head of the Large-Scale Assessment Research Group at ETS, which is responsible for the psychometric and statistical aspects of NAEP. James E. Carlson is responsible for psychometric and statistical analyses relating to special aspects and issues of NAEP.

- *U.S. History Framework for the 1994 National Assessment of Educational Progress* (NAGB, 1994).

- *Geography Framework for the 1994 National Assessment of Educational Progress* (NAGB, 1994).

- *NAEP 1994 Trends in Academic Progress: Achievement of U.S. Students in Science, 1969 to 1994; Mathematics, 1973 to 1994; Reading, 1971 to 1994; and Writing, 1984 to 1994* (Campbell, Reese, O'Sullivan, Dossey, & Donahue, 1996).

- *The 1994 NAEP Sampling and Weighting Report* (Wallace & Rust, 1996).

- *Report on Data Collection Activities for the 1994 National Assessment of Educational Progress* (Westat, Inc., 1995).

- *Report of Processing and Professional Activities* (National Computer Systems, 1995).

- *Technical Report of the NAEP 1994 Trial State Assessment Program in Reading* (Mazzeo, Allen, & Kline, 1995).

- *NAEP 1994 Trial State Assessment Program in Reading Secondary-Use Data Files User Guide* (O'Reilly, Zelenak, Rogers, & Kline, 1996a).

- *NAEP 1994 Secondary-Use Data Files User Guide* (O'Reilly, Zelenak, Rogers, & Kline, 1996b).

- *The NAEP Guide: A Description of the Content and Methods of the 1994 and 1996 Assessments* (NAEP, 1996).

The *First Look* reports highlight overall and demographic results, while the *Report Card* publications provide a more extensive treament of the findings and include discussions of contextual factors that are related to reading, U.S. history, and geography performance. For ordering information on these reports, write:

National Library of Education
Office of Educational Research and Improvement
U.S. Department of Education
555 New Jersey Avenue, NW
Washington, DC 20208-5641

or call 1-800-424-1616 (in Washington, DC metropolitan area call 202-219-1651).

The *Frameworks* are designed to assess the outcomes of students' education in reading, U.S. history, and geography in grade 4, 8, and 12 as part of NAEP. For ordering information on these reports, write:

National Assessment Governing Board
800 North Capitol Street NW
Suite 825
Washington, DC 20002

The *NAEP Guide* contains a description of the content and methods used in the 1994 and 1996 assessments; both the national and state components. It can be obtained by writing to:

National Library of Education
Office of Educational Research and Improvement
U.S. Department of Education
555 New Jersey Avenue, NW
Washington, DC 20208-5641

or call 1-800-424-1616 (in Washington, DC metropolitan area call 202-219-1651).

Additional samples of approximately 112,000 fourth-graders in public schools in 44 states and jurisdictions were assessed and another 10,000 fourth-grade students in nonpublic schools were assessed in reading as part of the 1994 Trial State Assessment. A representative sample of about 2,500 students was selected in each jurisdiction. The state-level sampling plan allowed for cross-state comparisons and comparisons with the nation in fourth-grade reading achievement. Technical details of the Trial State Assessment are not presented in this technical report but can be found in the *Technical Report of the NAEP 1994 Trial State Assessment Program in Reading* (Mazzeo, Allen, & Kline, 1995).

AN OVERVIEW OF NAEP IN 1994

For the 1994 assessment, NAEP researchers continued to build on the original design technology outlined in *A New Design for a New Era* (Messick, Beaton, & Lord, 1983). In order to maintain its links to the past and still implement innovations in measurement technology, NAEP continued its four-stage sampling approach. Long-term trend and short-term trend samples use the same methodology and population definitions as in previous assessments. Main assessment samples use innovations associated with new NAEP technology and address current educational issues. Long-term trend data are used to estimate changes in performance from previous assessments; main assessment sample data are used primarily for analyses involving the current student population, but also to estimate short-term trends for a small number of recent assessments. In continuing to use this two-tiered approach, NAEP reaffirms its commitment to maintaining long-term trends while at the same time implementing the latest in measurement technology.

In the 1994 assessment, many of the innovations that were implemented for the first time in 1988 were continued and enhanced. For example, a variant of the focused balanced incomplete block (focused-BIB) booklet design, which began in 1988 and was continued in 1992, was used in the 1994 main assessment samples in reading, U.S. history, and geography. In the focused-BIB design, an individual receives blocks of cognitive items in the same subject area. The focused-BIB design allows for improved estimation within a particular subject area, and estimation continues to be optimized for groups rather than individuals.

In 1994, NAEP continued to apply the plausible values approach to estimating means for demographic as well as curriculum-related subgroups. Proficiency estimates were based on "draws" from a posterior distribution that was based on an optimum weighting of two sets of information—the student's responses to cognitive items and his or her demographic and associated educational process variables. This Bayesian procedure was developed by Mislevy (see Chapter 11 or Mislevy, 1991). An improvement that was implemented first in 1988 and refined for the 1994 assessment is the multivariate procedure that uses information from all scales within a given subject area in the estimation of the proficiency distribution on any one scale in that subject area.

A major improvement in the 1992 and 1994 assessments was the use of the generalized partial credit model for IRT scaling. This allowed the incorporation of constructed-response questions that are scored on a multipoint rating scale into the NAEP scale in a way that utilizes the information available in each response category.

One important innovation in reporting the 1990 assessment data that was continued in 1992 and 1994 was the use of Bonferroni multiple comparison procedures in carrying out significance tests for the differences across assessment years. Methods such as the Bonferroni allow one to control for the type 1 error rate for a fixed number of comparisons. In addition to the Bonferroni procedures, tests for linear and quadratic trends were also applied to the national trend data in reading, mathematics, science, and writing. It is anticipated that future NAEP reports will continue to build on these developments, incorporating more powerful multiple comparison methods.

ORGANIZATION OF THE TECHNICAL REPORT

Part I of this report presents the details of the design of the 1994 National Assessment, summarized in Chapter 1. Chapters 2 through 8 describe the development of the objectives and the items used in the assessment, the sample selection procedures, the assessment booklets and questionnaires, the administration of the assessment in the field, the processing of the data from the assessment instruments into computer-readable form, the professional scoring of constructed-response items, and the methods used to create a complete NAEP database.

The 1994 NAEP data analysis procedures are described in Part II of the report. Chapter 9 provides a summary of the analysis steps. Subsequent chapters provide a general discussion of the weighting and variance estimation procedures used in NAEP, an overview of NAEP scaling methodology, and details of the trend and main assessment analyses performed for each subject area in the 1994 assessment.

Chapter 19 presents basic data from the 1994 assessment, including the properties of the measuring instruments and characteristics of the sample.

Chapter 1

OVERVIEW OF PART I: THE DESIGN AND IMPLEMENTATION OF THE 1994 NAEP[1]

Nancy L. Allen and Eugene G. Johnson
Educational Testing Service

1.1 INTRODUCTION

The 1994 National Assessment collected information on the knowledge, skills, understanding, and attitudes of young Americans in reading, U.S. history, geography, mathematics, science, and writing. The basis for this information was a complex sample survey involving more than 135,000 students, consisting of national samples of public- and nonpublic-school students who were aged 9, 13, and 17 or in grades 4, 8, and 12. Additional data came from the Trial State Assessment Program, which in 1994 assessed reading at grade 4 in representative samples of public-school students in 41 states, the District of Columbia, one Department of Defense Dependent School (International) jurisdiction and one territory. Some of the participating jurisdictions also tested representative samples of nonpublic-school students.

This chapter describes the design for the 1994 assessment and gives an overview of the steps involved in its implementation, from the planning stage through the creation of edited data files. The major components of the implementation process are presented here with references to the appropriate chapters in Part I for more details. The procedures used for the analysis of the data are summarized in the overview to Part II and discussed in detail in the remaining chapters in that part of the report. Excluded from this technical report are the details of the design and analysis of the 1994 Trial State Assessment, which instead appear in the *Technical Report of the NAEP 1994 Trial State Assessment Program in Reading* (Mazzeo, Allen, & Kline, 1995).

The organization of this chapter, and of Part I, is as follows:

- Section 1.2 provides an overview of the NAEP design for 1994 and describes the constituent samples. To provide background information, the section also gives the assessment schedule from the inception of NAEP in 1969 through the 1994 assessment.

- Section 1.3 summarizes the development of the objectives for each subject area in the assessment and describes the development and review of the items written to fit those objectives. Details of the objective and item development processes appear in Chapter 2.

---

[1] Nancy L. Allen is responsible for the psychometric and statistical analysis of national and state NAEP data. Eugene G. Johnson is a senior psychometrician with special expertise in the design of NAEP and sampling issues. The authors are indebted to the authors of Chapters 2 through 8 for portions of this chapter.

- Section 1.4 summarizes the four-stage stratified random sampling procedures used for the 1994 assessment with a fuller description provided in Chapter 3.
- Section 1.5 discusses the assignment of the cognitive and background questions to assessment booklets and describes the focused-BIB spiral design. Chapter 4 provides a detailed description of the assessment booklets.
- Section 1.6 summarizes the field administration procedures, including the processes of training field administrators, attaining school cooperation, administering the assessment, and conducting quality control. Further details appear in Chapter 5.
- Section 1.7 describes the flow of data from the receipt of the assessment materials through data entry, validation, and resolution to the creation of edited data files. Chapter 6 provides a detailed description of the process.
- Section 1.8 discusses the professional scoring of students' responses to the open-ended items in the assessment. Details of the process are given in Chapter 7.
- Section 1.9 summarizes the creation of the database, the quality control of data entry, and lists the 1994 database products. Further details appear in Chapter 8.

1.2 THE 1994 NAEP DESIGN

A major charge to NAEP is to reliably measure trends over time in educational achievement. To do this well, confounding effects due to changes from one assessment to the next in assessment instrumentation or in assessment procedures must be minimized. This implies a stability in the measurement process over time. At the same time, the assessment must remain current by allowing the introduction of new curriculum concepts and changes in educational priorities and by permitting the use of new measurement technology. The objectives for an assessment are determined through a consensus process in which committees of subject matter experts, scholars, and citizens representing many diverse constituencies and points of view are assembled to determine the educational goals that students should achieve. Satisfying these objectives often requires changes in assessment instrumentation and methodology.

A solution to the dilemma of measuring trends while maintaining currency is to institute a multicomponent assessment system where each component is itself an assessment designed to accomplish a specific goal. There are three components in the 1994 design: (1) assessments for long-term trend; (2) main assessments; and (3) the Trial State Assessment. These are discussed in detail below.

Several improvements were made in the design of NAEP in the 1984 and succeeding assessments. Until the 1984 assessment, NAEP was administered using matrix sampling and tape recorders; that is, by administering booklets of exercises using an aurally presented stimulus that paced groups of students through the individual assessment exercises in a common book. In the 1984 assessment, BIB spiraling, which does not include aural pacing, was introduced in place of taped matrix sampling. The NAEP design now includes sampling grade populations as well as the age populations that NAEP originally collected. The definitions of student age and the time of year in which the assessment takes place have been made uniform so that students in the fourth-, eighth-, and twelfth-grades are assessed. To shorten the timetable for reporting results, the period for national data collection was decreased in 1992 and 1994 from the five-month period used in 1990 to a three-month period in the winter (corresponding to the period used for the winter half-sample of the 1990 National Assessment). To enhance the coverage of the subject areas assessed, the number of cognitive items administered was increased for the 1992 and 1994 assessments.

The focus of assessments has also changed over time in response to changing priorities. For example, in contrast with previous assessments, the 1992 and 1994 assessments of reading required many more constructed responses and contained longer, more naturally occurring passages, including literary and informational texts as well as documents.

NAEP's design for 1994 required collecting 19 different samples in order to conduct the assessments. The various samples collected for the 1994 assessment are summarized in Table 1-1. Each row of Table 1-1 corresponds to a particular sample and each column of the table indicates the following major features of that sample:

1. *Sample* is the sample identifier. The first part of the sample code is a number representing the student cohort included in the sample; the second part, in brackets, denotes the specific sample type. For example, 9[Read-MainP] is a main assessment reading sample for age 9/grade 4. A full description of the purposes for the various sample types is given below.

2. *Booklets* gives the identifier numbers for the booklets used for the assessment of the particular sample.

3. *Mode* indicates the mode of assessment, which may be print, tape, or group. NAEP originally assessed students using a tape recorder in addition to booklets, thus pacing the students through exercises at a fixed rate. In 1994 NAEP used a paced audiotape for some of its mathematics and science long-term trend assessments. However, most other assessments in 1994 used printed instructions with the student expected to read the exercises. At grade 8, a national nonrepresentative sample of students participated in a pilot study in group assessment that was administered to collaborative groups of students, rather than individual students, who had participated in the U.S. history assessment.

4. The *cohort assessed* denotes the age, grade, or age/grade of the population being sampled. For example, *age 9/grade 4* represents students who are *either* 9 years old *or* in the fourth grade; an *age 17* cohort consists of students (in any grade) who are 17 years old; a *grade 8* cohort consists of students (of any age) who are in the eighth grade. The traditional NAEP samples used in long-term trend estimation were defined by age only. Populations for the 1994 main assessment and the 1994 reading and writing long-term trend assessments were defined by being either of a particular age *or* of the modal grade for students of that age. The definitions of age, and thus the corresponding grade, have changed in ways that are described in Section 1.2.1.

Table 1-1

NAEP 1994 Student Samples

Sample	Booklets	Mode	Cohort Assessed	Time of Testing	Age Definition	Modal Grade	Number Assessed
9[Rdg-MainP]	1 - 16	Print	Age 9/Grade 4	01/03/94 - 04/01/94	CY	4	9,680*e
13[Rdg-MainP]	1 - 18, 21	Print	Age 13/Grade 8	01/03/94 - 04/01/94	CY	8	13,547
17[Rdg-MainP]	1 - 18, 21, 22	Print	Age 17/Grade 12	01/03/94 - 04/01/94	CY	12	13,705
13[Rdg-MainR]*a	19, 20	Print	Age 13/Grade 8	01/03/94 - 04/01/94	CY	8	7,325
17[Rdg-MainR]	19, 20	Print	Age 17/Grade 12	01/03/94 - 04/01/94	CY	12	5,815
9[Geo-MainP]	31 - 48	Print	Age 9/Grade 4	01/03/94 - 04/01/94	CY	4	7,205
13[Geo-MainP]	31 - 49	Print	Age 13/Grade 8	01/03/94 - 04/01/94	CY	8	9,186
17[Geo-MainP]	31 - 49	Print	Age 17/Grade 12	01/03/94 - 04/01/94	CY	12	8,589
9[His-MainP]	101 - 118	Print	Age 9/Grade 4	01/03/94 - 04/01/94	CY	4	7,205
13[His-MainP]	101 - 133	Print	Age 13/Grade 8	01/03/94 - 04/01/94	CY	8	11,669
17[His-MainP]	101 - 133	Print	Age 17/Grade 12	01/03/94 - 04/01/94	CY	12	10,837
13[His-MainG]*b	134 - 135	Group	Grade 8	01/03/94 - 04/01/94	CY	8	257
9[Rdg-State]	1 - 16	Print	Grade 4	01/31/94 - 02/25/94	*c	4	*d
9[RW-LTTrend]	51 - 56	Print	Age 9/Grade 4	01/03/94 - 03/11/94	CY	4	5,335
13[RW-LTTrend]	51 - 56	Print	Age 13/Grade 8	10/11/93 - 12/17/93	CY	8	5,547
17[RW-LTTrend]	51 - 56	Print	Age 17/Grade 11	03/14/94 - 05/13/94	Not CY	11	4,840
9[MS-LTTrend]	91 - 93	Tape	Age 9	01/03/94 - 03/11/94	CY	4	5,663
13[MS-LTTrend]	91 - 93	Tape	Age 13	10/11/93 - 12/17/93	CY	8	6,052
17[MS-LTTrend]	84 - 85	Tape	Age 17	03/14/94 - 05/13/94	Not CY	11	3,813

a - **Note: The [Rdg-MainR] students participated in a study to examine how the choice of passages by students is related to their reading proficiency. This study is referred to as the NAEP Reader study.*

b - **Note: The 13[His-MainG] students participating in the pilot study of group assessment are a subsample of those in the 13[His-MainP] assessment.*

c - **Note: This sample was a "grade-only" sample.*

d - **Note: Consists of distinct samples in 44 jurisdictions.*

e - **Note: The smaller sample size for age 9/grade 4 is primarily due to the fact that the Reading to Perform a Task scale is not defined for this sample. Refer to Chapter 12 for more information.*

LEGEND

Rdg	*Reading*	*MainP*	*Main assessment, print administration*
Geo	*Geography*	*LTTrend*	*Long-term trend assessment*
His	*U.S. history*	*CY*	*Calendar year: birthdates in 1984, 1980, and 1976 respectively for ages 9, 13, and 17*
RW	*Reading and writing*		
MS	*Mathematics and science*	*Not CY*	*Age 17 only: birthdates between October 1, 1976, and September 30, 1977*
Print	*Print administration*		
Tape	*Audiotape administration*		

5. *Time of testing* indicates the time of year in which the assessment is performed. NAEP traditionally assessed 9-year-olds in the winter, 13-year-olds in the fall, and 17-year-olds in the spring; in the 1994 main assessment all age classes were assessed in the winter (between 1/3/94 and 4/1/94).

6. *Age definition* is denoted as calendar year (CY) or not calendar year (Not CY). NAEP originally defined age by birth within a calendar year at ages 9 and 13 but defined age 17 as being born between October 1 of one year and September 30 of the next. As in 1990 and 1992, the 1994 main assessments placed all ages on the same calendar-year basis.

7. The *modal grade* is the grade attended by most of the students of the sampled age. For example, if an age 17 sample is listed as having a modal grade of 11, then most of the 17-year-old students, as defined, are in the eleventh grade. The definition of age affects the modal grade of the sample. The ages in samples in the 1994 main assessment were defined so that the modal grades were 4, 8, and 12.

8. The *number assessed* is the number of students in the sample who were actually administered the assessment.

1.2.1 The 1994 NAEP Samples

The NAEP samples in 1994 consisted of three types—the main NAEP samples, the long-term trend samples, and the Trial State Assessment samples.

The Main NAEP Samples. The main NAEP samples are labeled in Table 1-1 as [Rdg-MainP], [Rdg-MainR], [His-MainP], [His-MainG], and [Geo-MainP]. The samples use the new assessment technology, focused-BIB spiraling (defined in Section 1.5) or special innovative procedures, and are intended to form the basis for future assessments. Each sample is assessed in the winter period (January 3 through April 1, 1994). In these samples, age is defined by calendar year, and both the age populations and modal grade populations are sampled. The main NAEP samples, and their purposes, are as follows:

> [Rdg-MainP] are age/grade reading assessment samples used for measuring reading achievement in 1994, as well as for measuring short-term trends in reading performance from 1992. The fourth-grade samples also provided the comparison groups for the 1994 Trial State Assessment of reading in grade 4. These samples used print administration.
>
> [Rdg-MainR] are age/grade reading assessments used to examine how the choice of passages by students is related to their reading proficiency. Some of these students selected one of seven stories to read, according to their interest; others were directed to read a particular story out of the seven in the NAEP Reader. NAEP Readers were not administered to age 9/grade 4 students. Results from the NAEP Reader samples are not reported in *NAEP 1994 Reading: A First Look—Findings from the National Assessment of Educational Progress* (Williams, Reese, Campbell, Mazzeo, & Phillips, 1995) nor in the *NAEP 1994 Reading Report Card for the Nation and the States: Findings from the National Assessment of Educational Progress*

and Trial State Assessments (Campbell, Donahue, Reese, & Phillips, 1996). Results from this special assessment are scheduled to be released as part of a focused report.

[His-MainP] are age/grade U.S. history assessment samples used for measuring U.S. history achievement in 1994. These samples were not used for comparison across assessment years since previous U.S. history assessments were based on different frameworks and objectives.

13[His-MainG] is a subsample of the students participating in the 13[His-MainP] assessment. These students participated in groups that completed sets of structured tasks. Group interactions were videotaped and analyzed by raters. Results from the group assessment are not reported in the *NAEP 1994 Geography Report Card: Findings from the National Assessment of Educational Progress* (Persky, Reese, O'Sullivan, Lazer, Moore, & Shakrani, 1996). The group assessment is described in more detail in Chapter 2.

[Geo-MainP] are age/grade geography assessment samples used for measuring world geography achievement. These samples were not used for short-term trend comparisons since previous geography assessments were based on different frameworks and objectives.

The Long-Term Trend Samples. The long-term trend samples are labeled as [RW-LTTrend] and [MS-LTTrend] in Table 1-1. Each sample was defined in the same way as equivalent samples in previous assessments and used the same assessment technology as was used in those assessments. Therefore, the long-term trend samples are directly comparable to data from previous assessments and so can be used for continuing the NAEP long-term trend lines. Because these samples were designed to link the 1994 data with data from previous assessments, they are also referred to as bridge samples. The long-term trend samples and their purposes are as follows:

[RW-LTTrend] are age/grade samples used for estimating long-term trends in reading and writing. These samples used assessment booklets identical to those initially used in 1984 and subsequently used in 1988, 1990, and 1992 (many of the items were also used in pre-1984 assessments). As in 1984, 1988, 1990, and 1992, print administration was used. These samples used the age definitions and time of testing originally used by NAEP in the 1970s and the early 1980s. The estimates of reading achievement from these samples link to seven previous reading assessments (1971, 1975, 1980, 1984, 1988, 1990, and 1992); the estimates of writing achievement link to four previous writing assessments (1984, 1988, 1990, and 1992).

[MS-LTTrend] are age-only samples used for estimating long-term trends in mathematics and science achievement. These samples used the same age definitions and time of testing as were used since 1969 and used the same assessment instruments as were used in the 1986, 1990, and 1992 long-term trend assessments of mathematics and science. As in previous assessments, the administration of the mathematics and science questions was paced with an audiotape. The estimates of science achievement from these samples link to seven previous science assessments (1970, 1973, 1977, 1982, 1986, 1990, and 1992); the estimates of mathematics achievement link to six previous assessments (1973, 1978, 1982, 1986, 1990, and 1992).

The Trial State Assessment Samples. In Table 1-1, 9[Rdg-State] are samples of fourth-grade public- and nonpublic-school students from each of the states and jurisdictions participating in the 1994 Trial State Assessment. The assessment booklets were the same print administered booklets as those used for the matching samples 9[Rdg-MainP], but the administrative procedures varied from that of the national assessment in that state personnel collected the data.

1.2.2 NAEP Assessments Since 1969

Table 1-2 shows the subject areas, grades, and ages assessed since the NAEP project began in 1969. As can be seen, in addition to the 1994 subject areas of reading, U.S. history, geography, mathematics, science, and writing, several other subject areas have been assessed over the years—social studies, civics, citizenship, literature, music, career development, art, and computer competence. Many subject areas are reassessed periodically to measure trends over time.

Assessments were conducted annually through 1980, but budget restrictions since then have reduced data collection to a biennial basis. Since its inception, NAEP has assessed 9-year-olds, 13-year-olds, and in-school 17-year-olds, although the age definitions changed in 1986 and again in 1988. Because of budget restrictions, NAEP no longer routinely assesses out-of-school 17-year-olds or young adults. (A separate assessment of young adults of ages 21 to 25 was conducted in 1985 under a separate grant.)

The table also indicates that in 1984, NAEP began gathering data by grade as well as by age, a practice that has been continued in assessments since then. It should be noted that somewhat different age definitions were used in the 1984, 1986, and 1988 assessments. In the 1984 assessment, the two younger ages were defined on a calendar-year basis, while the 17-year-olds were defined on an October 1 to September 30 basis. This resulted in modal grades of 4, 8, and 11. To allow for age cohorts that were exactly four years apart, in the 1986 main assessment all ages were defined on an October 1 to September 30 basis, resulting in modal grades of 3, 7, and 11. Special studies (Kaplan, Beaton, Johnson, & Johnson, 1988) were conducted to measure the effect of the changes in age definition. Because of problems encountered in assessing third graders, in 1988 the ages were redefined on a calendar-year basis, with the modal grades being 4, 8, and 12. These were the age definitions used in the 1990, 1992, and 1994 main assessments.

1.3 DEVELOPMENT OF ASSESSMENT OBJECTIVES, ITEMS, AND BACKGROUND QUESTIONS

In 1994, NAEP conducted main assessments of students at all three age/grade levels in reading, U.S. history, and geography. These assessments entailed the generation of a large number of cognitive items. In addition, a large number of background and attitude questions were asked of students and information was collected from principals and teachers. Details on the item development procedures for the 1994 main assessment are given in Chapter 2; this section provides an overview. (In addition to the main assessment, long-term trend studies were conducted in reading, mathematics, science, and writing. Since the instruments used for these studies consisted entirely of items used in previous assessments, no developmental tasks were required for their use in the 1994 assessment.)

In addition to the cognitive items, several questionnaires were developed: a common student background questionnaire given to all assessed students of a given age/grade, a school characteristics and policies questionnaire, teacher questionnaires for teachers of fourth- and eighth-grade students, and an IEP/LEP student questionnaire. Each of these questionnaires was developed through a broad-based consensus process.

Table 1-2
National Assessment of Educational Progress
Subject Areas, Grades, and Ages Assessed: 1969-1994

Assessment Year***	Subject Area(s)	Grades/Ages Assessed										
		Grade 3	Grade 4	Age 9	Grade 7	Grade 8	Age 13	Grade 11	Grade 12	Age 17	Age 17OS*	Adult
1969-70	Science			X			X			X	X	X
	Writing			X			X			X	X	X
	Citizenship			X			X			X	X	X
1970-71	Reading			X			X			X	X	X
	Literature			X			X			X	X	X
1971-72	Music			X			X			X	X	X
	Social Studies			X			X			X	X	X
1972-73	Science			X			X			X	X	X
	Mathematics			X			X			X	X	X
1973-74	Career and Occupational Development			X			X			X	X	X
	Writing			X			X			X	X	
1974-75	Reading			X			X			X	X	
	Art			X			X			X	X	
	Index of Basic Skills									X	X	

*__Note__: *Age 17 students who had dropped out of school or had graduated prior to assessment.*

**__Note__: *Small, special-interest assessment conducted on limited samples at specific grades or ages.*

***__Note__: *It should be noted that somewhat different age definitions were used in the 1984, 1986, and 1988 assessments. In the 1984 assessment, the two younger ages were defined on a calendar-year basis, while the 17-year-olds were defined on an October 1 to September 30 basis. This resulted in modal grades of 4, 8, and 11. To allow for age cohorts that were exactly four years apart, in the 1986 main assessment all ages were defined on an October 1 to September 30 basis, resulting in modal grades of 3, 7, and 11. Special studies (Kaplan, Beaton, Johnson, & Johnson, 1988) were conducted to measure the effect of the changes in age definition. Because of problems encountered in assessin third graders, in 1988 the ages were redefined on a calendar-year basis, with the modal grades being 4, 8, and 12. These were the age definitions used in the 1990, 1992, and 1994 main assessments.*

Table 1-2 (continued)

National Assessment of Educational Progress

Subject Areas, Grades, and Ages Assessed: 1969-1994

Assessment Year***	Subject Area(s)	Grades/Ages Assessed										
		Grade 3	Grade 4	Age 9	Grade 7	Grade 8	Age 13	Grade 11	Grade 12	Age 17	Age 17OS*	Adult
1975-76	Citizenship/Social Studies			X			X			X	X	
	Mathematics**						X			X	X	
1976-77	Science			X			X			X		
	Basic Life Skills**									X		
	Science, Reading, Energy, Health**											X
1977-78	Mathematics			X			X			X		
	Consumer Skills**									X		
1978-79	Writing, Art, and Music			X			X			X		
1979-80	Reading/Literature			X			X			X	X	
	Art						X					
1981-82	Science**			X			X			X		
	Mathematics and Citizenship/Social Studies			X			X			X		

__Note__: Age 17 students who had dropped out of school or had graduated prior to assessment.

**__Note__: *Small, special-interest assessment conducted on limited samples at specific grades or ages.*

***__Note__: *It should be noted that somewhat different age definitions were used in the 1984, 1986, and 1988 assessments. In the 1984 assessment, the two younger ages were defined on a calendar-year basis, while the 17-year-olds were defined on an October 1 to September 30 basis. This resulted in modal grades of 4, 8, and 11. To allow for age cohorts that were exactly four years apart, in the 1986 main assessment all ages were defined on an October 1 to September 30 basis, resulting in modal grades of 3, 7, and 11. Special studies (Kaplan, Beaton, Johnson, & Johnson, 1988) were conducted to measure the effect of the changes in age definition. Because of problems encountered in assessing third graders, in 1988 the ages were redefined on a calendar-year basis, with the modal grades being 4, 8, and 12. These were the age definitions used in the 1990, 1992, and 1994 main assessments.*

Table 1-2 (continued)

National Assessment of Educational Progress
Subject Areas, Grades, and Ages Assessed: 1969-1994

Assessment Year***	Subject Area(s)	Grades/Ages Assessed										
		Grade 3	Grade 4	Age 9	Grade 7	Grade 8	Age 13	Grade 11	Grade 12	Age 17	Age 17OS*	Adult
1984	Reading		X	X		X	X	X		X		
	Writing		X	X		X	X	X		X		
	Writing (long-term trend)		X	X		X	X	X		X		
1985	Adult Literacy**											X
1986	Reading	X		X	X		X	X		X		
	Mathematics	X		X	X		X	X		X		
	Science	X		X	X		X	X		X		
	Computer Competence	X		X	X		X	X		X		
	U.S. History**							X		X		
	Literature**							X		X		
	Reading (long-term trend)		X	X		X	X	X		X		
	Mathematics (long-term trend)		X	X		X	X	X		X		
	Science (long-term trend)		X	X		X	X	X		X		

*__Note__: *Age 17 students who had dropped out of school or had graduated prior to assessment.*

**__Note__: *Small, special-interest assessment conducted on limited samples at specific grades or ages.*

***__Note__: *It should be noted that somewhat different age definitions were used in the 1984, 1986, and 1988 assessments. In the 1984 assessment, the two younger ages were defined on a calendar-year basis, while the 17-year-olds were defined on an October 1 to September 30 basis. This resulted in modal grades of 4, 8, and 11. To allow for age cohorts that were exactly four years apart, in the 1986 main assessment all ages were defined on an October 1 to September 30 basis, resulting in modal grades of 3, 7, and 11. Special studies (Kaplan, Beaton, Johnson, & Johnson, 1988) were conducted to measure the effect of the changes in age definition. Because of problems encountered in assessing third graders, in 1988 the ages were redefined on a calendar-year basis, with the modal grades being 4, 8, and 12. These were the age definitions used in the 1990, 1992, and 1994 main assessments.*

Table 1-2 (continued)

National Assessment of Educational Progress
Subject Areas, Grades, and Ages Assessed: 1969-1994

Assessment Year***	Subject Area(s)	Grades/Ages Assessed										
		Grade 3	Grade 4	Age 9	Grade 7	Grade 8	Age 13	Grade 11	Grade 12	Age 17	Age 17OS*	Adult
1988	Reading		X	X		X	X		X	X		
	Writing		X	X		X	X		X	X		
	Civics		X	X		X	X		X	X		
	U.S. History		X	X		X	X		X	X		
	Document Literacy**					X	X		X	X		
	Geography**								X	X		
	Reading (long-term trend)		X	X		X	X	X		X		
	Writing (long-term trend)		X	X		X	X	X		X		
	Mathematics (long-term trend)			X			X	X		X		
	Science (long-term trend)			X			X	X		X		
	Civics (long-term trend)						X			X		

***Note:** *Age 17 students who had dropped out of school or had graduated prior to assessment.*

****Note:** *Small, special-interest assessment conducted on limited samples at specific grades or ages.*

*****Note:** *It should be noted that somewhat different age definitions were used in the 1984, 1986, and 1988 assessments. In the 1984 assessment, the two younger ages were defined on a calendar-year basis, while the 17-year-olds were defined on an October 1 to September 30 basis. This resulted in modal grades of 4, 8, and 11. To allow for age cohorts that were exactly four years apart, in the 1986 main assessment all ages were defined on an October 1 to September 30 basis, resulting in modal grades of 3, 7, and 11. Special studies (Kaplan, Beaton, Johnson, & Johnson, 1988) were conducted to measure the effect of the changes in age definition. Because of problems encountered in assessing third graders, in 1988 the ages were redefined on a calendar-year basis, with the modal grades being 4, 8, and 12. These were the age definitions used in the 1990, 1992, and 1994 main assessments.*

Table 1-2 (continued)
National Assessment of Educational Progress
Subject Areas, Grades, and Ages Assessed: 1969-1994

Assessment Year***	Subject Area(s)	Grades/Ages Assessed										
		Grade 3	Grade 4	Age 9	Grade 7	Grade 8	Age 13	Grade 11	Grade 12	Age 17	Age 17OS*	Adult
1990	Reading		X	X		X	X		X	X		
	Mathematics		X	X		X	X		X	X		
	Science		X	X		X	X		X	X		
	Reading (long-term trend)		X	X		X	X	X		X		
	Writing (long-term trend)		X	X		X	X	X		X		
	Mathematics (long-term trend)			X			X			X		
	Science (long-term trend)			X			X			X		
	Trial State Mathematics					X						
1992	Reading		X	X		X	X		X	X		
	Writing		X	X		X	X		X	X		
	Mathematics		X	X		X	X		X	X		
	Reading (long-term trend)		X	X		X	X	X		X		
	Writing (long-term trend)		X	X		X	X	X		X		
	Mathematics (long-term trend)			X			X			X		
	Science (long-term trend)			X			X			X		
	Trial State Mathematics		X			X						
	Trial State Reading		X									

*__Note__: *Age 17 students who had dropped out of school or had graduated prior to assessment.*

**__Note__: *Small, special-interest assessment conducted on limited samples at specific grades or ages.*

***__Note__: *It should be noted that somewhat different age definitions were used in the 1984, 1986, and 1988 assessments. In the 1984 assessment, the two younger ages were defined on a calendar-year basis, while the 17-year-olds were defined on an October 1 to September 30 basis. This resulted in modal grades of 4, 8, and 11. To allow for age cohorts that were exactly four years apart, in the 1986 main assessment all ages were defined on an October 1 to September 30 basis, resulting in modal grades of 3, 7, and 11. Special studies (Kaplan, Beaton, Johnson, & Johnson, 1988) were conducted to measure the effect of the changes in age definition. Because of problems encountered in assessing third graders, in 1988 the ages were redefined on a calendar-year basis, with the modal grades being 4, 8, and 12. These were the age definitions used in the 1990, 1992, and 1994 main assessments.*

Table 1-2 (continued)

National Assessment of Educational Progress
Subject Areas, Grades, and Ages Assessed: 1969-1994

Assessment Year***	Subject Area(s)	Grades/Ages Assessed										
		Grade 3	Grade 4	Age 9	Grade 7	Grade 8	Age 13	Grade 11	Grade 12	Age 17	Age 17OS*	Adult
1994	Reading		X	X		X	X		X	X		
	U.S. History		X	X		X	X		X	X		
	Geography		X	X		X	X		X	X		
	Reading (long-term trend)		X	X		X	X	X		X		
	Writing (long-term trend)		X	X		X	X	X		X		
	Mathematics (long-term trend)			X			X			X		
	Science (long-term trend)			X			X			X		
	Trial State Reading		X									

*__Note__: *Age 17 students who had dropped out of school or had graduated prior to assessment.*

**__Note__: *Small, special-interest assessment conducted on limited samples at specific grades or ages.*

***__Note__: *It should be noted that somewhat different age definitions were used in the 1984, 1986, and 1988 assessments. In the 1984 assessment, the two younger ages were defined on a calendar-year basis, while the 17-year-olds were defined on an October 1 to September 30 basis. This resulted in modal grades of 4, 8, and 11. To allow for age cohorts that were exactly four years apart, in the 1986 main assessment all ages were defined on an October 1 to September 30 basis, resulting in modal grades of 3, 7, and 11. Special studies (Kaplan, Beaton, Johnson, & Johnson, 1988) were conducted to measure the effect of the changes in age definition. Because of problems encountered in assessing third graders, in 1988 the ages were redefined on a calendar-year basis, with the modal grades being 4, 8, and 12. These were the age definitions used in the 1990, 1992, and 1994 main assessments.*

All items in the assessment underwent extensive reviews by subject area and measurement specialists, as well as careful scrutiny to eliminate any potential bias or lack of sensitivity to any group. Further, the items were field tested on a representative group of students. Based on the results of the field test, items were revised or modified as necessary and then again reviewed for bias. With the help of staff and outside reviewers, the Instrument Development Panels selected the items to include in the assessment. After the items were selected and formed into the final groupings or blocks of items, they were carefully reviewed by the National Center for Education Statistics (NCES), the Office of Management and Budget (OMB), and the National Assessment Governing Board (NAGB).

The assessment instruments included multiple-choice, short constructed-response, and extended constructed-response questions in every subject area. The constructed-response questions were professionally scored as described in Chapter 7.

1.4 THE 1994 SAMPLE DESIGN

The sample for the 1994 NAEP assessment was selected using a complex four-stage sample design involving the sampling of students from selected schools within 94 selected geographic areas, called primary sampling units (PSUs), across the United States. The sample design was similar to that used in 1986, 1988, 1990, and 1992 and is described in detail by Westat, Inc., the firm subcontracted by ETS to select the sample, in *1994 National Assessment of Educational Progress Sampling and Weighting Procedures, Final Report* (Wallace & Rust, 1996). The following sections provide an overview of each of the four stages of the sampling design with further details given in Chapter 3.

➪ Stage 1: Primary Sampling Units

In the first stage of sampling, the United States (the 50 states and the District of Columbia) was divided into geographic PSUs. Each PSU met a minimum size requirement and generally comprised either a metropolitan statistical area (MSA), a single county, or a group of contiguous counties. The PSUs were classified into four regions (Northeast, Southeast, Central, West), each containing about one-fourth of the U.S. population. In each region, PSUs were additionally classified as MSA or nonMSA. This resulted in eight subuniverses of PSUs.

Ninety-four of the PSUs were selected for the 1994 main assessment. Twenty-two PSUs were designated as certainty units because of their size, as it was cost effective to include them in the sample with certainty. The remaining smaller PSUs were not guaranteed to be selected and were accordingly designated as noncertainty PSUs. Within each major stratum (subuniverse), further stratification was achieved by ordering the noncertainty PSUs according to several additional socioeconomic characteristics. Seventy-two PSUs were selected, one per stratum from each of the noncertainty strata, with probability proportional to size (the number of school-age children from the 1990 census). To enlarge the samples of Black and Hispanic students, thereby enhancing the reliability of estimates for these groups, PSUs from the high-minority subuniverses were sampled at twice the rate of those from the other subuniverses. This was achieved by creating smaller strata within the high-minority subuniverses.

For the long-term trend samples, 52 PSUs were selected: Ten PSUs were selected with certainty; six additional PSUs were selected from the 12 remaining main sample certainty PSUs; and 36 PSUs were selected from the 72 noncertainty strata independently of PSU selections for the main samples.

Stage 2: Sampling Schools

In the second stage of sampling for the main assessments, the public schools (including Bureau of Indian Affairs schools and Department of Defense schools) and nonpublic schools (including Catholic schools) within each of the selected PSUs were listed according to the three age/grades. An independent sample of schools was selected separately for each of the age/grades so that some schools were selected for assessment of two age/grades, and a few were selected for all three. Schools that had been selected for long-term trend at a particular age (9, 13, or 17) were not selected for the main assessment at the same age. Schools within each PSU were selected (without replacement) with probabilities proportional to assigned measures of size with oversampling of nonpublic schools and of schools with high minority enrollment. Overall probabilities of selection for high-minority schools were twice those for other schools while the probabilities of selection of nonpublic schools were triple those of low-minority public schools of the same size. The increased probabilities of selection enlarged the samples of Black and Hispanic students and the samples of students from nonpublic schools, thereby enhancing the reliability of estimates for these groups. Details of the probabilities used for school selection appear in Chapter 3.

The samples of schools for the long-term trend assessments were drawn in a manner very similar to that used for the main assessments. The chief difference in the two samples was that nonpublic schools and schools with high minority enrollment were not oversampled for these assessments. Schools were not selected for both main and long-term trend assessments at the same age/grade.

For the main samples, the overall school cooperation rate was 86 percent for the two lower age/grade levels; it was 79 percent for age 17/grade 12. For the long-term trend samples, the overall school cooperation rate was 87 percent for age class 9, 82 percent for age class 13, and 81 percent for age class 17. In certain instances, refusing schools were replaced by substitutes according to the rules indicated in Chapter 3.

Stage 3: Assigning Assessment Sessions to Schools

In the third stage of sampling, assessment sessions were assigned to the sampled schools, as described in Chapter 3. An assessment session typically consisted of 25 to 30 students, all of whom could be assessed following the same procedures. There were two general types of sessions in the 1994 assessment: (1) tape sessions, where every student was administered the same booklet and where audiotape prompts paced the students through at least part of the booklet, and (2) print sessions, where a number of distinct booklets were administered and where no audiotape pacing was used. (Print sessions are also called spiral sessions, since the assessment booklets were spiraled for administration—see Section 1.5.1.) The assignment of sessions to schools was designed to maximize the number of session types conducted within each PSU, where each session type corresponded to a separate sample of the population of students.

Stage 4: Sampling Students

In the fourth stage of sampling, a consolidated list was prepared for each school of all grade-eligible and age-eligible students for the age class for which the school was selected. To provide the target sample size, a systematic selection of eligible students was made from this list, if necessary. In small and medium-sized schools all eligible students were in the sample. For schools assigned to more than a single session type, students were assigned by Westat district supervisors to one of the various session types using specified procedures (audiotape administration, print administration, or group interaction). No student was assigned to more than one session, except in the case of the U.S. history pilot study of group assessment.

➪ Stage 4a: Excluded Students

Despite NAEP's goal to assess all selected students, certain selected students were judged by school authorities as being incapable of participating meaningfully in the assessment. For each of these students, school staff who have knowledge of students' capabilities completed an excluded student questionnaire, listing the reason for exclusion and providing some background information.

Specific guidelines for exclusion were provided for all samples in the 1994 assessment. However, somewhat different criteria were used for the long-term trend samples than for the main assessment samples. The exclusion guidelines for the long-term trend samples were the same as those used in previous assessments. Three types of students could be excluded under these guidelines—non-English speaking students, students with mental retardation who are educable but who were judged incapable of meaningfully responding to exercises appropriate to their age level, and students so functionally disabled that they could not perform in the NAEP assessment situation.

The criteria used for the main assessments since 1990 provided more specific rules for exclusion. A student identified as having Limited English Proficiency (LEP) could be excluded if the student was a native speaker of a language other than English, had been enrolled in an English-speaking school for less than two years, and was judged incapable of taking part in the assessment. Students to be excluded for reasons other than LEP had to be special education students with Individualized Education Plans (IEP), or equivalent designation, who were mainstreamed less than 50 percent of the time in academic subjects and/or were considered unassessable by the IEP team.

➪ Stage 4b: Sampling Teachers

The reading teachers of fourth- and eighth-grade students participating in the main assessment of reading, the U.S. history teachers of fourth- and eighth-grade students participating in the main assessment of U.S. history, and the geography instructors of fourth- and eighth-grade students participating in the main assessment of geography were identified and asked to complete a questionnaire (described in Chapter 2) about their background and experiences and about instructional practices, by class, for any classes containing assessed students.

➪ Stage 4c: The School Characteristics and Policies Questionnaires

A school characteristics and policies questionnaire was mailed to every sampled school by Westat before the assessment for completion by the principal or school administrator. The Westat supervisor then collected the questionnaires and returned them to ETS. The school characteristics and policies questionnaire is described in Chapter 2.

1.5 ASSESSMENT INSTRUMENTS

Four types of instruments were used in the 1994 assessment: student assessment booklets (which included the student common background questionnaire), IEP/LEP student questionnaires, teacher questionnaires, and a school characteristics and policies questionnaire. This section provides an overview of these instruments; more detailed information can be found in Chapter 4.

1.5.1 Student Assessment Booklets—Main Assessment

The student assessment booklets for the 1994 assessment contained both cognitive and noncognitive questions. Each assessment booklet contained common background questions, subject-specific background questions, and either one or two blocks of cognitive questions.

As in previous assessments, the assembly of most cognitive items into booklets and their subsequent assignment to assessed students was determined by a *focused balanced incomplete block* (focused-BIB) design with *spiraled* administration. The first step in implementing BIB spiraling is to divide the items within a subject area into units called blocks, where each block is designed to take 25 or 50 minutes to complete. (The long-term assessments of reading, writing, mathematics, and science used previously created 15-minute blocks of items—the details of the long-term assessment booklets are given later.)

The 25-minute blocks were assembled into booklets containing two blocks of subject area items according to a balanced incomplete block design. In addition, background questions were included in the booklets. In a balanced incomplete block design, the subject area blocks are assigned to booklets in such a way that each block appears in the same number of booklets and every pair of blocks appears together in at least one booklet. This is the *balanced* part of the method; the *incomplete* part refers to the fact that no booklet contains all items and hence incomplete data is yielded for each assessed student.

As in 1988, 1990, and 1992, this BIB design was *focused*—that is, each block of items within each of the subject areas was paired with other blocks within that subject area but not with blocks of items from other subject areas.

> **Reading.** Unlike the design for the U.S. history and geography assessments, the reading items were assembled using a *partially balanced incomplete block* (PBIB) design.
>
> The instruments for the 1994 assessment of reading for age 9/grade 4 included eight 25-minute cognitive blocks of items while the instruments for each of the two older age/grades included nine 25-minute blocks plus two to three 50-minute blocks of items. Each 25-minute block contained one relatively long reading passage and 8 to 12 constructed-response and multiple-choice questions. At the two older age/grades, three of the nine 25-minute blocks focused on each of three purposes for reading: Reading for Literary Experience, Reading for Information, and Reading to Perform a Task.
>
> The three literary experience blocks (L_1, L_2, L_3,), the three information blocks (I_1, I_2, I_3), and the three task-oriented blocks (T_1, T_2, T_3) at each of age 13/grade 8 and age 17/grade 12 were paired into 18 booklets according to the design shown in Table 1-3.

This design is partially balanced in the sense that every block within a given purpose for reading is paired with every other block measuring the same purpose but is only paired with one of the three available blocks measuring one of the other two purposes (a fully balanced design in which all 9 blocks are paired together would require an additional 18 booklets). In this design, every block appears in exactly four booklets, twice in booklets measuring a single purpose for reading and twice in booklets measuring two purposes. The order of administration is balanced, with each block appearing twice as the first cognitive block in the booklet and twice as the second. Furthermore, to partially balance context effects in the booklets measuring two of the purposes for reading, each purpose appears first in at least one booklet and second in at least one other.

Table 1-3
Main Assessment Reading Booklet Design
Age 13/Grade 8 and Age 17/Grade 12

Booklet	Blocks
1	L_1 L_2
2	L_2 L_3
3	L_3 L_1
4	I_1 I_2
5	I_2 I_3
6	I_3 I_1
7	T_1 T_2
8	T_2 T_3
9	T_3 T_1
10	L_1 I_2
11	I_3 L_2
12	L_3 I_1
13	I_1 T_2
14	I_2 T_3
15	T_1 I_3
16	L_2 T_1
17	T_2 L_3
18	T_3 L_1

Table 1-4

Main Assessment Reading Booklet Design
Age 9/Grade 4

Booklet	Blocks
1	L_1 L_2
2	L_2 L_3
3	L_3 L_4
4	L_4 L_1
5	L_1 L_3
6	L_2 L_4
7	I_1 I_2
8	I_2 I_3
9	I_3 I_4
10	I_4 I_1
11	I_1 I_3
12	I_2 I_4
13	I_3 L_1
14	I_4 L_2
15	L_3 I_1
16	L_4 I_2

At age 9/grade 4, four of the blocks focused on Reading for Literary Experience and four more focused on Reading for Information. For age 9/grade 4, the four literary experience blocks (L_1, L_2, L_3, L_4) and the four information blocks (I_1, L_2, I_3, I_4) were paired into 16 booklets as shown in Table 1-4.

Like the design for the two older age/grades, this design is partially balanced in that every block within a given purpose for reading is paired with every other block measuring the same purpose but is only paired with one block measuring the other of the two purposes. Every block appears in four booklets, three times within booklets measuring the same purpose and once in a booklet measuring both purposes. The order of administration is balanced with each block appearing twice in the first position and twice in the second position of a booklet. Order effects in the four booklets containing both purposes are balanced since each purpose appears first in two booklets and second in the remaining two.

Each booklet at each age/grade also contained five minutes of a common core of background questions and five minutes of general and reading-related background questions. Students in the fourth grade were given additional time because the items in the general questionnaire were read aloud to them. A one-minute questionnaire was also given to students at the end of each booklet to determine students' motivation in completing the assessment and their familiarity with assessment tasks. At age 13/grade 8 and age 17/grade 12, additional booklets were created, each consisting of the common and reading-related background questions and a single block requiring 50 minutes for completion. One booklet was created for age 13/grade 8, focusing on Reading for Information. For age 17/grade 12, two additional booklets were created focusing on Reading for Information. In addition, one booklet at age 13/grade 8 and one at age

17/grade 12 consisted of background questions and a single 50-minute block, and was used in the *NAEP Reader* study.

U.S. History. The instruments for the 1994 assessment of U.S. history included six 25-minute blocks of items at age 9/grade 4, and eight 25-minute blocks of items at the higher two age/grade levels. For each of the two higher age/grade levels, one booklet contained a 50-minute block of items. Items from any of the four major U.S. history themes (Democracy, Cultures, Technology, and World Role) may be found in each 25-minute block. The 50-minute blocks were constructed with the intent of assessing one theme thoroughly, and most, if not all, of the items from a specific 50-minute block were from one of the themes. The six 25-minute blocks administered to age 9/grade 4 were paired into 18 booklets according to the BIB design shown in Table 4-7 in Chapter 4. Each block of items appears as the first cognitive block and as the second cognitive block three times in the design, controlling context and fatigue effects. The eight 25-minute blocks administered to age 13/grade 8 and age 17/grade 12 were paired into 32 booklets according the BIB design in Table 4-9 in Chapter 4. Each block of items appears as the first cognitive block and as the second cognitive block four times in this design. Each booklet also contained five minutes of a common core of background questions and five minutes of background questions related to U.S. history. Students in the fourth grade were given additional time because the items in the general questionnaire were read aloud to them. A one-minute questionnaire was also given to students at the end of each booklet to determine students' motivation in completing the assessment and their familiarity with assessment tasks.

Geography. The instruments for the 1994 assessment of geography included six 25-minute blocks of items at each age/grade level. For each of the two higher age/grade levels, one booklet contained a 50-minute block of items. Items from any of the three content areas, Space and Place, Environment and Society, and Spatial Dynamics and Connections, may be found in each block. The six 25-minute blocks administered to each age/grade were paired into 18 booklets according to the same design used for the age 9/grade 4 U.S. history assessment (see Tables 4-12 and 4-14 in Chapter 4). As in the U.S. history six-block design, each cognitive geography block appears in each position exactly three times. As for reading and U.S. history, each geography booklet also contained five minutes of common background questions and five minutes of subject-related background questions. Students in the fourth grade were given additional time because the items in the general questionnaire were read aloud to them. A one-minute questionnaire was also given to students at the end of each booklet to determine students' motivation in completing the assessment and their familiarity with assessment tasks.

For the main assessment, 52 different booklets were assembled for age 9/grade 4, 72 were assembled for age 13/grade 8, and 73 were assembled for age 17/grade 12. Each booklet consisted of cognitive items from a single subject area. Within each subject area, certain cognitive items were presented to two or three ages.

These booklets were spiraled within subject area and type of administration (paced or print) and placed into bundles. Spiraling involves interleaving the booklets in regular (systematic) sequence so that each booklet appears an appropriate number of times in the sample. Booklets were packaged together in bundles large enough to accommodate a typical assessment session. The bundles were designed so that each booklet would appear equally often in each position in a bundle.

The final step in the BIB-spiraling procedure was the assigning of the booklets to the assessed students. The students within an assessment session were assigned booklets in the order in which the booklets were bundled. As a result, typically, each student in an assessment session received a different booklet and, even in schools with multiple sessions, only a few students received the same booklet or block of items. In the 1994 BIB-spiral design, representative and randomly equivalent national samples of about 2,350 grade- or age-eligible students responded to each item (resulting in samples of about 2,000 students eligible by age and 2,000 eligible by grade).

The BIB-spiral design permits the estimation of correlations between all items within a content area and the estimation of correlations of estimates of proficiency between content areas within a subject area. Furthermore, since the spiral design presents each block of items to fewer persons in any school, but to more schools, than would a simpler matrix sampling design, the cluster effect is markedly reduced, leading to a sample with high statistical efficiency.

A further benefit of the BIB-spiral design is that it balances the position of items across booklets. In a simple matrix sampling scheme, each item appears in only one position in a booklet. In particular, the same items always occur last in a booklet, resulting in potential underestimates of student ability because of fatigue factors. In the BIB-spiral design, each block of items occurs once in each block position—for instance, first and second for the two-block designs. For the two-block BIB design, half of the students administered an item were presented the item in the front half of a booklet; the other half were presented the item in the latter half of a booklet. As a result of this balancing, NAEP has found that in assessments that rely more heavily on reading or on constructed responses, students generally do less well on the last block of questions in a booklet, regardless of the content. However, also as a result of this balancing, NAEP results are not unduly influenced by the fact that students tend to perform less well on items occurring at the end of booklets.

The spiral design does preclude the use of audiotape pacing. Since each student within a session responds to a different set of items, audiotaped administration would be unmanageable. The instructions and the items themselves must be read by the student. For this reason, simple matrix sampling designs, where every student in an assessment session receives the same booklet, are used in the special cases requiring audiotape pacing, such as the long-term trend assessments of mathematics and science.

1.5.2 Student Assessment Booklets—Long-Term Trend Samples

There were two distinct long-term trend samples in the 1994 assessment, each of which required reprinting booklets used in previous assessments:

Reading-Writing Long-Term Trend: Six booklets were used at each of the three age/grades for the purposes of measuring long-term trends in reading and writing. These booklets were identical to booklets used in the 1984 main assessments of reading and writing and in the 1988, 1990, and 1992 long-term trend assessments of those subjects. Each booklet consisted of a common background block and three cognitive blocks, either two reading and one writing or one reading and two writing. All cognitive blocks also contained subject-related background questions. The booklets were administered without audiotape and were spiraled together for administration.

***Mathematics-Science Long-Term Trend*:** These instruments were used for the measurement of mathematics and science and were identical to booklets administered in 1990 and 1992. These booklets contained 15-minute blocks of mathematics and science items; each mathematics block and each science block was administered using audiotape pacing. (At the younger two ages, the booklets also contain a block of reading items, which was print-administered.) There were three booklets each at age 9 and age 13 and two booklets at age 17. Combined, the booklets at an age contain three blocks of mathematics items and three blocks of science items. Because of the audiotape pacing, each booklet was administered in a separate session.

1.5.3 Other Instruments

Besides the student assessment booklets, other instruments provided data relating to the assessment:

The ***IEP/LEP Student Questionnaires*** were completed by the teachers of those students who were selected to participate in the assessment sample who had Individualized Education Plans (IEP) or were classified as Limited English Proficient (LEP). The questionnaires were completed for all IEP or LEP students, whether or not they actually participated in the assessment. The questionnaires asked about the nature of the student's disability and the special programs in which the student participated. The criteria used for excluding students are described in Chapter 5.

Teacher questionnaires were administered to the teachers of all students participating in the fourth- and eighth-grade reading, U.S. history, and geography assessments. The teacher questionnaire included a general section that contained questions about the teacher's background and experience. The rest of the questionnaire contained questions about instructional practices, by class, for any classes containing assessed students.

School characteristics and policies questionnaires were completed by school principals or their representatives, who provided information about school administration, staffing patterns, special programs, subject requirements, and school resources.

1.6 FIELD OPERATIONS AND DATA COLLECTION

Field operations and data collection for the 1994 assessment were the responsibility of Westat, Inc., and are documented in Chapter 5. The field operation was conducted by a staff at Westat's home office and a larger staff in the field. The Westat home office staff coordinated all activities related to field operations and managed materials distribution and home-office receipt of assessment reporting forms. The field staff consisted of area supervisor, assessment supervisors, and exercise administrators. The assessment supervisors, who were trained by Westat, were each responsible for the assessment activities in one or more PSUs. Although ETS made initial contact with participating school districts, each assessment supervisor was primarily responsible for making follow-up contacts with these districts, recruiting and training exercise administrators to work with them in administering the assessment sessions, arranging the assessment sessions, and selecting the sample of students to be assessed within each school. The assessment supervisors and the exercise administrators administered the assessments, filled out the necessary forms, performed process control, and shipped the assessment booklets and forms to National Computer Systems (NCS), the subcontractor responsible for processing NAEP materials and data.

Gaining school cooperation was the joint responsibility of Westat and ETS staff. ETS made the preliminary contacts preparatory to obtaining school cooperation by first contacting the Chief State School Officers, informing them that schools within their states had been selected for the assessment and, in a later letter, listing the selected schools and districts. Later mailings were sent to superintendents of public schools and parochial schools and principals of other nonpublic schools for all schools selected in the assessment. These materials provided an explanation of NAEP, a list of the selected schools in the official's jurisdiction, and a cover letter explaining that a Westat district supervisor would contact them to set up an introductory meeting. Westat district supervisors then scheduled and conducted introductory meetings, worked with the schools to schedule the assessments, and, with the exercise administrators, conducted the assessments. The overall cooperation rate of schools originally selected was 83 percent for the 1994 main samples and for the long-term trend samples. Further detail on school participation rates is given in Chapter 3.

The main assessment sessions were conducted between January 3 and April 11, 1994, at all age/grade levels. The age 9/grade 4 long-term trend assessments were carried out between January 3 and March 11; the age 17/grade 11 long-term trend samples were conducted between March 14 and May 13, 1994. The age 13/grade 8 long-term trend assessments were carried out between October 11 and December 17, 1993. When the main assessments of the long-term trend subjects were first collected in 1987 and 1986, studies were completed to take into account the difference in assessment time across the samples (Kaplan, Beaton, Johnson, & Johnson, 1988).

An automated management system tracked and recorded the progress of field work throughout the 1994 assessment period. In addition, progress was constantly monitored through telephone reports held between the area supervisors and the assessment supervisors and between the area supervisors and the home office staff.

Both Westat and ETS participated in the quality control of the field administration, which involved on-site visits by Westat and ETS staff to verify the sampling of the students and to observe the conduct of the assessment by the supervisors and the exercise administrators.

1.7 MATERIALS AND DATA PROCESSING

After completing an assessment session, Westat field supervisors and exercise administrators shipped the assessment booklets and forms from the field to National Computer Systems for entry into computer files, professional scoring, and creating the data files for transmittal to ETS. Careful checking assured that all data from the field were received. More than 215,000 booklets and questionnaires were received and processed for the national portion of the 1994 assessment. The extensive processing of these data is detailed in Chapter 6.

The student data were transcribed into machine-readable form by scanning the student instruments with an optical scanning machine. An intelligent data entry system was used for resolution of the scanned data, the entry of documents rejected by the scanning machine, and the entry of information from the questionnaires. Additionally, each piece of input data was checked to verify that it was of an acceptable type, that it was within a specified range or ranges of values, and that it was consistent with other data values. The entry and editing of materials is discussed in Chapter 6.

1.8 PROFESSIONAL SCORING

Items requiring a written response from the student (constructed-response items) were included in the main assessments in reading, U.S. history, and geography, the Trial State Assessment in reading, and the long-term trend assessments in reading, mathematics, science, and writing. More than 2.3 million constructed responses were read and marked by the professional scoring staff for the national portion of the 1994 assessment.

Image processing and scoring were used in NAEP for the first time in 1994. Images of students' responses to the constructed-response items were scanned into computerized form, then scored online by professional raters.

Chapter 7 describes the professional scoring operation, including an overview of the scoring guides, the training procedures, and the scoring process for each subject area.

1.9 CREATION OF THE DATABASE

Before analyses could begin, the student response data, school, teacher, and IEP/LEP student questionnaire data, and all sampling weights had to be integrated into a coherent and comprehensive database. This database, which was used for all analyses, was also the source for the creation of two NAEP database products—the item information database and the secondary-use data files. Secondary-use data files include sample control statement files for SAS and SPSS statistical systems and the NAEP Data on Disk product suite. The Data on Disk products, including a complete set of secondary-use data files on CD-ROM, PC-based NAEP data extraction software, and NAEP analysis modules, make secondary use of NAEP data much easier than it has been in the past. The quality of the data resulting from the complete data entry system, from the actual instruments collected in the field to the final machine-readable database used in analysis, was verified by selecting field instruments at random and performing a character-by-character comparison of these instruments with their representations in the final database. Chapter 8 provides details on the database, quality control activities, and database products.

Chapter 2

DEVELOPING THE NAEP OBJECTIVES, ITEMS, AND BACKGROUND QUESTIONS FOR THE 1994 ASSESSMENTS OF READING, U.S. HISTORY, AND GEOGRAPHY[1]

Stephen Lazer, Jay R. Campbell, and Patricia Donahue
Educational Testing Service

2.1 INTRODUCTION

The subject areas constituting the main 1994 assessments were reading, U.S. history, and geography.[2] Long-term trend assessments were also conducted in reading, writing, mathematics, and science; these assessments used items that were developed as part of previous assessments. Additional data was gathered under the auspices of the Trial State Assessment Program, which in 1994 assessed reading at grade 4 in representative samples of public-school students in 41 participating states, the District of Columbia, Guam, and at Department of Defense Education Activity Overseas Schools.[3] In addition, some of the participating states also tested representative samples of nonpublic-school students. A summary of each main assessment subject area follows:

> **Reading:** A national reading assessment was administered at grades 4, 8, and 12. This assessment was designed around questions requiring in-depth analysis of authentic, naturally occurring reading materials. A mixture of multiple-choice, short constructed-response, and extended constructed-response questions made up the assessment; in aggregate, well over half of the student assessment time was spent answering constructed-response rather than multiple-choice questions. In addition, a special study entitled "The NAEP Reader" was conducted at grades 8 and 12, in which students were given a choice of reading material, and the effects of this on performance were studied. For the participants in the Trial State Assessment Program, the entire grade 4 assessment was administered.
>
> **U.S. History:** At grades 4, 8, and 12, a newly developed U.S. history assessment was conducted. This assessment was built around exercises that required students to interpret a broad variety of primary and secondary U.S. history materials, and to show knowledge of a broad range of history content. The assessment included both multiple-choice and constructed-response questions, and students spent, on average, more than half their time answering the latter.

---

[1] Stephen Lazer manages assessment development activities for the NAEP program at ETS. He also coordinated the development of the 1994 U.S. history and geography assessments. Jay R. Campbell is the reading assessment coordinator for NAEP and Patricia Donahue participates in the development of the reading assessments and reports.

[2] Copies of the frameworks for these assessments are available from the National Assessment Governing Board (NAGB).

[3] Two of the states, Idaho and Michigan, did not meet minimum participation rate guidelines for data reporting. One jurisdiction, Washington, D.C., withdrew from the Trial State Assessment after the data collection phase.

Geography: A newly developed assessment at grades 4, 8, and 12, composed entirely of multiple-choice and constructed-response tasks, was administered to nationally representative samples of students. A number of constructed-response questions asked students to create maps, charts, and tables, while others required them to write responses. A wide variety of stimulus materials—including maps, photographs, atlases, tables, charts, graphs, and cartoons—were used at each grade. Students were required to demonstrate both content knowledge and analytic skills.

From its inception, NAEP has developed assessments through a consensus process and the 1994 instruments were no exception. Under the direction of the National Assessment Governing Board (NAGB), educators, scholars, and citizens representative of many diverse constituencies and points of view designed an assessment framework for each of the three subject areas, proposing goals they felt students should achieve in the course of their education. Staff at Educational Testing Service (ETS) who were subject-area experts in their respective fields worked with subject-area consultants well versed in assessment methodology to develop assessment questions appropriate to the objectives. All questions underwent extensive reviews by subject-matter specialists, measurement specialists, and ETS employees trained to conduct "sensitivity" reviews designed to prevent the inclusion in assessments of materials that might prove offensive or unfair to subpopulations in American society. Questions were assembled and printed into booklets suitable for matrix sampling and then administered either by a trained field staff (for the national program) or by state or local school district staff (for the Trial State Assessment Program) to stratified, multistage probability samples of students.

All 1994 development efforts were governed by four major considerations:

1. As specified in NAEP's authorizing legislation, the objectives were to be developed through a consensus process involving subject-matter experts, school administrators, teachers, and parents, and the items were to be reviewed carefully for potential bias.

2. As outlined in the ETS proposal for the administration of the NAEP contract, the development of questions for each subject area was to be guided by an Instrument Development Committee that contained considerable overlap with the panelists involved in guiding the framework consensus process.

3. As described in the *ETS Standards for Quality and Fairness* (ETS, 1987), all materials developed at ETS were to be in compliance with specified procedures. In particular, all questions were to be carefully reviewed for content accuracy, testworthiness, and potential bias.

4. As required by federal regulations, all NAEP items were to be submitted to a complex clearance process. This process involved review of all cognitive items by NCES and NAGB, and review of all background questions by the Office of Management and Budget (OMB), the Information Management and Compliance Division (IMCD) of the Department of Education, and NCES.

The development effort for the 1994 assessment included questionnaires[4] for students, teachers, and school administrators, in addition to a substantial number of cognitive items for each of the three subject areas.

The following sections include general overviews about setting objectives and developing items and specific details about developing subject-specific objectives and assessments. A list of the consultants who participated in the 1994 development process is included in Appendix A.

2.2 GENERAL OVERVIEW OF THE 1994 ASSESSMENT FRAMEWORKS

The subject-area objectives for each NAEP assessment are determined through a legislatively mandated consensus process. These objectives typically take the form of frameworks, or matrices, delineating the important content and process areas to be assessed. The various frameworks for the 1994 assessments are described below.

All the frameworks for the main 1994 NAEP assessments were developed through consensus processes conducted by the Council of Chief State School Officers (CCSSO), who were working under contract to NAGB. The projects involved participation and review by many groups, including teachers, content-area scholars, educational policymakers, and members of the general public. In addition to people directly involved in the framework development processes, the documents were reviewed by state education and testing officials, by representatives of professional associations, and by researchers. In addition, the frameworks were the subject of testimony at public hearings arranged to allow the widest possible participation in the consensus process. The objectives resulting from these processes reflect neither a narrowly defined theoretical framework nor the view of every participant. They do, however, represent the thinking of a broad cross-section of individuals who are expert in the various content areas and who are deeply committed to the improvement of American education.

The **reading** assessment framework used in both 1992 and 1994 is a four-by-three matrix specifying four reading stances and three reading purposes. The stances are Initial Understanding, Developing an Interpretation, Personal Reflection and Response, and Demonstrating a Critical Stance. The three global reading purposes are Reading for Literary Experience, Reading to Gain Information, and Reading to Perform a Task. The 1994 reading assessment was designed to measure students' abilities to read based on a variety of passages, ranging from informational materials, documents, and news articles to poems, essays, and stories. The stimuli used in NAEP represent the types of materials that students commonly encounter in and out of school and are expected to be capable of reading.

NAEP's 1994 **U.S. history** assessment was based on a new framework that focused assessment developers on the measurement of four history themes. These themes are:

1. *Democracy* - Change and Continuity in American Democracy: Ideas, Institutions, Practices, and Controversies;

2. *Cultures* - The Gathering and Interactions of Peoples, Cultures, and Ideas;

---

[4] These questionnaires can be obtained from the National Center for Education Statistics (NCES).

3. *Technology* - Economic and Technological Changes and Their Relationship to Society, Ideas, and the Environment; and

4. *World Role* - The Changing Role of America in the World.

In addition to these themes, the U.S. history framework specifies eight history periods and two cognitive dimensions for measurement. These periods and cognitive dimensions will be described in depth below.

The 1994 U.S. history assessment was designed to measure both knowledge of history and history reasoning skills. To accomplish these measurement goals, the framework called for the use of a wide variety of stimulus materials (including primary documents, photographs, cartoons, artworks, graphs, and tables). It also required in-depth analysis of a single U.S. history topic for some components and that no more than half of the allotted assessment time be spent on multiple-choice items.

The 1994 **geography** framework is a three-by-three matrix specifying three content areas and three cognitive dimensions. The content areas are Space and Place, Environment and Society, and Spatial Dynamics and Connections. The cognitive dimensions are Knowing, Understanding, and Applying. To ensure well-rounded coverage of the field of geography, the framework instructed that a variety of exercise-types—including those that required students to create their own maps or to work with atlases—be used in the assessment. As was the case in U.S. history, no more than half of the allotted assessment time could be spent on multiple-choice items.

2.3 GENERAL OVERVIEW OF PROCEDURES FOR DEVELOPING THE ITEMS

A carefully developed and tested series of steps, similar to those used for past NAEP assessments, was utilized to create assessment items that reflected the reading, U.S. history, and geography objectives and that measured achievement related to them (see Sections 2.4—2.6 for more detail). The steps were as follows:

1. Item specifications and prototype items were provided in conjunction with the frameworks in each subject area.

2. The Instrument Development Committees in each subject area provided guidance to NAEP staff about ways in which the objectives could be measured given the realistic constraints of resources and measurement technology. The committees made recommendations about priorities for the assessments (within the context of the assessment framework) and the types of questions to be developed.

3. In reading, the existing pool of items was reviewed in detail and trend blocks (that is, blocks administered in precisely the same form as in the 1992 NAEP assessment and used to measure change in student proficiencies) were selected.

4. Specialists with subject-matter expertise, skills, and experience in creating items according to specifications were identified from inside and outside ETS to develop and review the assessment questions. The development of items and related materials (passages, stimuli, etc.) therefore involved the input of practitioners from around the country as well as from members of the Instrument Development Committees.

5. Newly created questions were reviewed and revised by staff and external reviewers, including the Instrument Development Committee. The items for the fourth-grade reading assessment were also reviewed by representatives from the State Education Agencies. Also for the reading assessment, blocks of materials and questions were sent to teachers across the country for review.

6. Further language editing and sensitivity reviews were conducted as required by the *ETS Standards for Quality and Fairness*.

7. Field test materials were prepared, including the materials necessary to secure OMB clearance. The clearance packages were also reviewed by representatives of NCES, and NAGB had responsibility for approving the cognitive materials.

8. Field tests for the national program were conducted with representative groups of students from across the country.

9. Field-test booklets were scored and the results analyzed to examine both item level statistical performance and, in a limited way, differential item functioning.

10. Based on these analyses and the results of the field test, questions were deleted, revised, or modified and re-edited. They once again went through the required ETS sensitivity review.

11. With the help of staff and outside reviewers, each Instrument Development Committee selected the items or, in the case of reading, the blocks to include in the operational assessments.

12. In U.S. history and geography, where blocks did not move from field test to operational assessment in intact form, field-tested items were assembled into operational form blocks with attention given to balancing content coverage and difficulty levels.

13. Each block underwent final content, editorial, and sensitivity reviews.

14. After a final check to ensure that each assessment booklet and each block therein met the overall guidelines for the assessment, the materials were sent for final OMB clearance, NCES review, and NAGB approval for cognitive materials.

15. The booklets were typeset and printed.

The following sections describe the development of the reading, U.S. history, and geography assessments in more detail.

2.4 DEVELOPING THE READING ASSESSMENT

2.4.1 Overview

The framework that was developed for the 1992 NAEP assessment in reading also served as the framework for the 1994 assessment. Similar to all previous NAEP assessments, the objectives in reading were developed through a broad-based consensus process. To prepare the framework and objective for the 1994 assessment, NAGB contracted with CCSSO. The development process involved a steering committee, a planning committee, and CCSSO project staff. Educators, scholars, and citizens, representative of many diverse constituencies and points of view, participated in the national consensus process to design objectives for the reading assessment.

The instrument used in the 1994 reading assessment was composed of a combination of reading passages and questions from the 1992 assessment and a set of passages and questions newly developed for 1994. Those passages and questions carried over from the 1992 instrument comprised two-thirds of the 1994 instrument. The remaining third was made up of new passages and questions developed according to the framework. Maintaining two-thirds of the instrument across the two assessment years allowed for the reporting of trends in reading performance. At the same time, developing a new set of passages and questions made it possible to release one-third of the 1992 assessment for public use.

In developing the new portion of the 1994 NAEP reading assessment, the same framework, objectives, and procedures used in 1992 were followed. After careful review of the objectives, reading materials were selected and questions were developed that were appropriate to the objectives. All questions were extensively reviewed by specialists in reading, measurement, and bias/sensitivity, as well as by state representatives. The core goals of the ETS assessment development process and procedures used to realize these goals are outlined in the introduction to this chapter and in Section 2.3.

Sections 2.4.2 through 2.4.4 include a detailed description of the development of the framework, objectives, and items for the 1994 NAEP reading assessment. Section 2.4.8 also describes the student background questionnaires and the reading teacher questionnaire. Additional information on the structure and content of assessment booklets can be found in Chapter 4. Various committees worked on the development of the framework, objectives, and items for the reading assessment. The list of committee members and consultants who participated in the 1994 development process is provided in Appendix A.

2.4.2 Development of the Assessment Framework

NAGB is responsible for setting policy for NAEP; this policymaking role includes the development of assessment frameworks and test specifications. Appointed by the Secretary of Education from lists of nominees proposed by the Board itself in various statutory categories, the 24-member board is composed of state, local, and federal officials, as well as educators and members of the public.

The NAGB began the development process for the 1992 reading objectives (which also served as the objectives for the 1994 assessment) by conducting a widespread mail review of the objectives for the 1990 reading assessment and by holding a series of public hearings throughout the country. The contract for managing the remainder of the consensus process was awarded to the CCSSO. The development process included the following activities:

- A Steering Committee consisting of members recommended by each of 15 national organizations (see Appendix A) was established to provide guidance for the consensus process. The committee monitored the progress of the project and offered advice. Drafts of each version of the document were sent to members of the committee for review and reaction.

- A Planning Committee (see Appendix A) was established to identify the objectives to be assessed in reading and prepare the framework document. The members of this committee consisted of experts in reading, including college professors, an academic dean, a classroom teacher, a school administrator, state level assessment and reading specialists, and a representative of the business community. This committee met with the Steering Committee and as a separate group. A subgroup also met to develop item specifications. Between meetings, members of the committee provided information and reactions to drafts of the framework.

- The project staff at CCSSO met regularly with staff from NAGB and NCES to discuss progress made by the Steering and Planning committees.

During this development process, input and reactions were continually sought from a wide range of members of the reading field, experts in assessment, school administrators, and state staff in reading assessment, in particular, innovative state assessment efforts and work being done by the Center for the Learning and Teaching of Literature (Langer, 1989, 1990).

2.4.3 Framework and Assessment Design Principles

The reading objectives framework was designed to focus on reading processes and outcomes, rather than reflect a particular instructional or theoretical approach. It was stated that the framework should focus not on the specific reading skills that lead to outcomes, but rather on the quality of the outcomes themselves. The framework was intended to embody a broad view of reading by addressing the increasing level of literacy needed for employability, personal development, and citizenship. The framework also specified a reliance on contemporary reading research and the use of nontraditional assessment formats that more closely resemble desired classroom activities.

The objectives development was guided by the consideration that the assessment should reflect many of the curricular emphases and objectives in various states, localities, and school districts in addition to what various scholars, practitioners, and interested citizens believed should be included in the curriculum. Accordingly, the committee gave attention to several frames of reference:

- The purpose of the NAEP reading assessment is to provide information about the progress and achievement of students in general rather than to test individual students' ability. NAEP is designed to inform policymakers and the public about reading ability in the United States.

- The term "reading literacy" should be used in the broad sense of knowing when to read, how to read, and how to reflect on what has been read. It represents a complex, interactive process that goes beyond basic or functional literacy.

- The reading assessment should use valid and authentic tasks that are both broad and complete in their coverage of important reading behaviors so that the test will be useful and valid, and will demonstrate a close link to desired classroom instruction.
- Every effort should be made to make the best use of available methodology and resources in driving assessment capabilities forward. New types of items and new methods of analysis were recommended for NAEP reading assessments.
- Every effort must be made in developing the assessment to represent a variety of opinions, perspectives, and emphases among professionals, as well as state and local school districts.

2.4.4 Framework for the 1994 Assessment

The framework adopted for the 1992 reading assessment—and used again in 1994—was organized according to a four-by-three matrix of reading *stances* by reading *purposes*. The stances include

- Initial Understanding,
- Developing an Interpretation,
- Personal Reflection and Response, and
- Demonstrating a Critical Stance.

These stances were assessed across three global purposes defined as

- Reading for Literary Experience,
- Reading to Gain Information, and
- Reading to Perform a Task.

Different types of texts were used to assess the various purposes for reading. Students' reading abilities were evaluated in terms of a single purpose for each type of text. At grade 4, only Reading for Literary Experience and Reading to Gain Information were assessed, while all three global purposes were assessed at grades 8 and 12. Figure 2-1 and 2-2 describe the four reading stances and three reading purposes that guided the development of NAEP's 1992 and 1994 reading assessments.

Figure 2-1

Description of Reading Stances

Readers interact with text in various ways as they use background knowledge and understanding of text to construct, extend, and examine meaning. The NAEP reading assessment framework specified four reading stances to be assessed that represent various interactions between readers and texts. These stances are not meant to describe a hierarchy of skills or abilities. Rather, they are intended to describe behaviors that readers at all developmental levels should exhibit.

Initial Understanding

Initial understanding requires a broad, preliminary construction of an understanding of the text. Questions testing this aspect ask the reader to provide an initial impression or unreflected understanding of what was read. The first question following a passage was usually one testing initial understanding.

Developing an Interpretation

Developing an interpretation requires the reader to go beyond the initial impression to develop a more complete understanding of what was read. Questions testing this aspect require a more specific understanding of the text and involve linking information across parts of the text as well as focusing on specific information.

Personal Reflection and Response

Personal reflection and response requires the reader to connect knowledge from the text more extensively with his or her own personal background knowledge and experience. The focus is on how the text relates to personal experience; questions on this aspect ask the readers to reflect and respond from a personal perspective. Personal reflection and response questions were typically formatted as constructed-response items to allow for individual possibilities and varied responses.

Demonstrating a Critical Stance

Demonstrating a critical stance requires the reader to stand apart from the text, consider it, and judge it objectively. Questions on this aspect require the reader to perform a variety of tasks such as critical evaluation, comparing and contrasting, application to practical tasks, and understanding the impact of such text features as irony, humor, and organization. These questions focus on the reader as critic and require reflection on and judgments about how the text is written.

Figure 2-2

Description of Purposes for Reading

Reading involves an interaction between a specific type of text or written material and a reader, who typically has a purpose for reading that is related to the type of text and the context of the reading situation. The reading assessment presented three types of text to students representing each of three reading purposes: literary text for literary experience, informational text to gain information, and documents to perform a task. Students' reading skills were evaluated in terms of a single purpose for each type of text.

Reading for Literary Experience

Reading for literary experience involves reading literary text to explore the human condition, to relate narrative events with personal experiences, and to consider the interplay in the selection among emotions, events, and possibilities. Students in the NAEP reading assessment were provided with a wide variety of literary text, such as short stories, poems, fables, historical fiction, science fiction, and mysteries.

Reading to Gain Information

Reading to gain information involves reading informative passages in order to obtain some general or specific information. This often requires a more utilitarian approach to reading that requires the use of certain reading/thinking strategies different from those used for other purposes. In addition, reading to gain information often involves reading and interpreting adjunct aids such as charts, graphs, maps, and tables that provide supplemental or tangential data. Informational passages in the NAEP reading assessment included biographies, science articles, encyclopedia entries, primary and secondary historical accounts, and newspaper editorials.

Reading to Perform a Task

Reading to perform a task involves reading various types of materials for the purpose of applying the information or directions in completing a specific task. The reader's purpose for gaining meaning extends beyond understanding the text to include the accomplishment of a certain activity. Documents requiring students in the NAEP reading assessment to perform a task included directions for creating a time capsule, a bus schedule, a tax form, and instructions on how to write a letter to a senator. Reading to perform a task was assessed only at grades 8 and 12.

The Planning Committee was interested in creating an assessment that would be forward-thinking and reflect quality instruction. In recognition that the demands made of readers change as they mature and move through school, it was recommended that the proportion of items have some relation to reading purpose (for literary experience, to gain information, to perform a task,). The distribution of items by reading purpose across grade levels recommended in the assessment framework is provided in Table 2-1.

Table 2-1
Percentage Distribution of Items by Reading Purpose as Specified in the NAEP Reading Framework

Grade	Purpose for Reading		
	Reading for Literary Experience	**Reading to Gain Information**	**Reading to Perform a Task**
4	55%	45%	(Not Assessed)
8	40%	40%	20%
12	35%	45%	20%

Readers use a range of cognitive abilities and assume various stances that should be assessed within each of the reading purposes. While reading, students form an initial understanding of the text and connect ideas within the text to generate interpretations. In addition, they extend and elaborate their understanding by responding to the text personally and critically and by relating ideas in the text to prior experiences or knowledge. Table 2-2 shows the distribution of items by reading stance, as specified in the reading framework, for all three grade levels.

Table 2-2
Percentage Distribution of Items by Reading Stance as Specified in the NAEP Reading Framework

Grade	Reading Stance		
	Initial Understanding/ Developing an Interpretation	**Personal Reflection and Response**	**Demonstrating a Critical Stance**
4, 8, and 12	33%	33%	33%

2.4.5 Developing the Cognitive Items

The development of cognitive items began with a careful selection of grade-appropriate passages for the assessment. Passages were selected from a pool of reading selections contributed by teachers from across the country. The framework stated that the assessment passages should represent authentic, naturally occurring reading material that students may encounter in and out of school. Furthermore, these passages were to be reproduced in test booklets as they had appeared in their original publications. In some cases, materials (such as bus schedules) were provided to students separate from the printed assessment booklet. Final passage selections were made by the Reading Instrument Development Committee. In order to guide the development of items, passages were outlined or mapped to identify essential elements of the text.

The assessment included constructed-response (short and extended) and multiple-choice items. The decision to use a specific item type was based on a consideration of the most appropriate format for assessing the particular objective. Both types of constructed-response items were designed to provide an in-depth view of students' ability to read thoughtfully and to respond appropriately to what they read. Short constructed-response questions were used when students needed to respond in only one or two sentences in order to demonstrate full comprehension. Extended constructed-response questions were used when the task required more thoughtful consideration of the text and engagement in more complex reading processes. Multiple-choice items were used whenever a reading outcome could be measured through use of these items.

A carefully developed and proven series of steps was used to create the assessment items. These steps are described earlier in this chapter under Section 2.3.

The assessment included 25-minute and 50-minute "blocks," each consisting of one or more passages and a set of multiple-choice and constructed-response items to assess students' comprehension of the written material. Students were asked to respond to either two 25-minute blocks or one 50-minute block (at grades 8 and 12). The grade 4 assessment included eight 25-minute blocks (four blocks measuring each of the two global purposes for reading assessed at this grade). The instruments at grades 8 and 12 each included nine 25-minute blocks (three blocks measuring each of the global purposes for reading). In addition, the grade 8 assessment included two 50-minute blocks and the grade 12 assessment included three 50-minute blocks.

One of the 50-minute blocks at grades 8 and 12 was the *NAEP Reader*, which asked students to choose a story to read and to answer questions about that story. A related study, in which a pre-specified number of students was asked to read specific stories in the reader. It allowed for the examination of the impact of choice on student performance.

2.4.6 Development of the Operational Forms

The 1992 reading assessment was reviewed to determine which blocks of exercises should be included in the 1994 assessment for trend purposes. In all, two-thirds of the 1992 assessment were carried forward for use in 1994.

Reading field tests were conducted in February and March 1993 and involved national samples of fourth-, eighth-, and twelfth-grade students. The intent of the field test was to try out the items and to allow for the selection of replacement blocks for those that had been released. In addition, the field test allowed NAEP to introduce to the assessment constructed-response questions that could be scored for partial credit. At grade 4, 32 items were field tested; the grade 8 field test included 81 items; and the grade 12 field test included 80 items. These figures may tend to understate the field test effort, since items for the *NAEP Reader* were counted only once but were analyzed and scored separately depending on students' choices of story. About 500 responses were obtained for each item in the field test.

The field test data were collected, scored, and analyzed in preparation for meetings with the Reading Instrument Development Committee. Using item analysis, which provided the mean percentage of correct responses, the polyserial correlations, and the difficulty level for each item in the field test, committee members, ETS test development staff, and NAEP/ETS staff reviewed the materials. The objectives that guided these reviews included

- determining which items were most related to overall student achievement,
- determining the need for revisions of items that lacked clarity or had ineffective item formats,
- prioritizing items to be included in the assessment, and
- determining appropriate timing for assessment items.

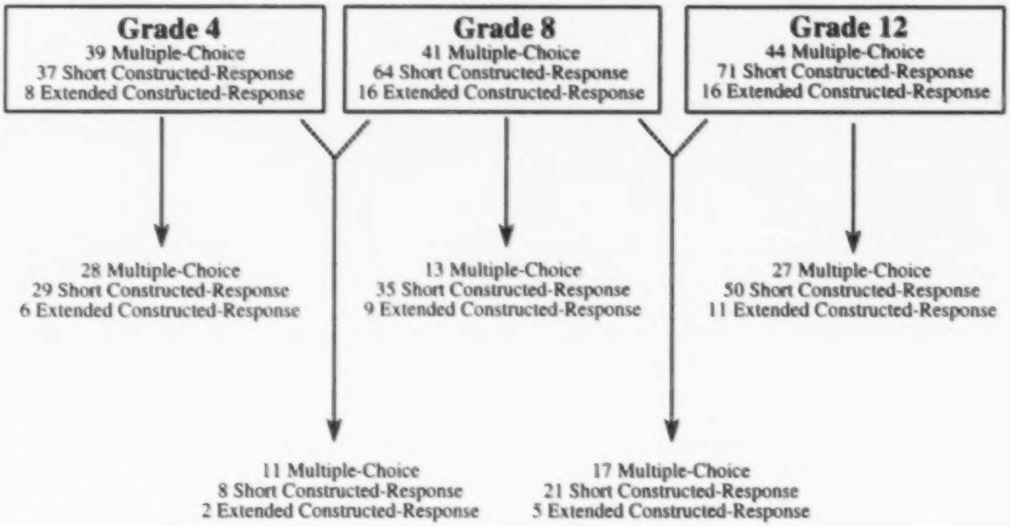
Once the committees had selected the items, all items were rechecked for content, measurement, and sensitivity concerns. The federal clearance process was initiated in June 1992 with the submission of draft materials to NCES. The package containing the final set of cognitive items assembled into blocks and questionnaires was submitted in June 1993. Throughout the clearance process, revisions were made in accordance with changes required by the government. Upon approval, the blocks (assembled into booklets) and questionnaires were ready for printing.

2.4.7 Distribution of Assessment Items

Table 2-3 lists the total number of items at each grade level in the 1994 assessment. Of the total of 272 items, there are 96 unique multiple-choice items and 176 unique constructed-response questions that make up the 1994 reading assessment. Some of these items are used at more than one grade level. As a result, the sum of the items that appear at each grade level is greater than the total number of unique items.

Table 2-3

Distribution of Items for the 1994 Reading Assessment


Grade 4
39 Multiple-Choice
37 Short Constructed-Response
8 Extended Constructed-Response
Grade 8
41 Multiple-Choice
64 Short Constructed-Response
16 Extended Constructed-Response
Grade 12
44 Multiple-Choice
71 Short Constructed-Response
16 Extended Constructed-Response
28 Multiple-Choice
29 Short Constructed-Response
6 Extended Constructed-Response
13 Multiple-Choice
35 Short Constructed-Response
9 Extended Constructed-Response
27 Multiple-Choice
50 Short Constructed-Response
11 Extended Constructed-Response
11 Multiple-Choice
8 Short Constructed-Response
2 Extended Constructed-Response
17 Multiple-Choice
21 Short Constructed-Response
5 Extended Constructed-Response

In the development process, every effort was made to meet the content and process targets specified in the assessment framework. Table 2-4 shows the approximate percentage of aggregate assessment time devoted to each purpose for reading, at each grade level. Percentages are based on the classifications agreed upon by NAEP's 1994 Instrument Development Committee. Note that the numbers presented in Table 2-4 differ from Table 2-1 in that Table 2-1 shows the distribution of assessment time as specified in the reading framework.

Table 2-4

Percent Distribution of Assessment Time by Grade and Reading Purpose

Reading Purpose	Grade 4	Grade 8	Grade 12
Reading for Literary Experience	50%	38%	33%
Reading to Gain Information	50%	38%	47%
Reading to Perform a Task	N/A	23%	20%

Table 2-5 shows the approximate percentage of assessment time devoted to each reading stance. Unlike the purposes for reading, in which individual students did not receive questions in all areas, every student completed tasks involving each of the reading stances. It is recognized that making discrete classifications is difficult for these categories and that independent efforts to classify NAEP questions have led to different results (National Academy of Education, 1992). Also, it has been found that developing personal response questions that are considered equitable across students' different backgrounds and experiences is difficult. Note that the numbers presented in Table 2-5 differ from Table 2-2 in that Table 2-2 shows the distribution of items as specified in the reading framework.

Table 2-5

*Percentage Distribution of Assessment Time by Grade and Reading Stance**

Reading Stance	Grade 4	Grade 8	Grade 12
Initial Understanding/ Developing an Interpretation	49%	50%	46%
Personal Reflection and Response	25%	20%	19%
Demonstrating a Critical Stance	27%	30%	35%

**Note: Percentages may not total 100 due to rounding.*

2.4.8 Reading Background Questionnaires

Research indicates that school, home, and attitudinal variables affect students' reading comprehension and literacy. Therefore, in addition to assessing how well students read, it is important to understand the instructional context in which reading takes place, students' home support for literacy, and their reading habits and attitudes. To gather contextual information, NAEP assessments include background questions designed to provide insight into the factors that may influence reading proficiency in the literary, informational, and document categories assessed.

NAEP includes both general background questionnaires given to participants in all subjects and subject-specific questionnaires for both students and their teachers. The development of the general background questionnaires is discussed below. It is worth noting that members of the Reading Instrument Development Committee were consulted on the appropriateness of the issues addressed in all questionnaires that may relate to reading instruction and achievement. Like the cognitive items, all background questions were submitted for extensive review and field testing. Recognizing the validity problems inherent in self-reported data, particular attention was given to developing questions that were meaningful and unambiguous and that would encourage accurate reporting.

In addition to the cognitive questions, the 1994 assessment included two five-minute sets of general and reading background questions designed to gather contextual information about students, their instructional and recreational experiences in reading, and their attitudes toward reading. Students in the fourth grade were given additional time because the items in the general questionnaire were read aloud for them. A one-minute questionnaire was also given to students at the end of each booklet to determine students' motivation in completing the assessment and their familiarity with assessment tasks.

➪ Student Reading Questionnaires

Three categories of information were represented in the five-minute sections of reading background questions called the **student reading questionnaire** (15 questions at grade 4, 25 questions at grade 8, and 26 questions at grade 12):

> *Time Spent Studying Reading*: Students were asked to describe the amount of instruction they received in reading and the time spent on reading homework.
>
> *Instructional Practice*: Students were asked to report their instructional experiences related to reading in the classroom, including group work, special projects, and writing in response to reading. In addition, they were asked about the instructional practices of their reading teachers and the extent to which the students themselves discussed what they read in class and demonstrated use of skills and strategies.
>
> *Attitudes Towards Reading*: Students were asked a series of questions about their attitudes and perceptions about reading, such as whether they enjoyed reading and whether they were good in reading.

➪ Reading Teacher Questionnaire

To supplement the information on instruction reported by students, the reading teachers of the fourth and eighth graders participating in the NAEP reading assessment were asked to complete a questionnaire about their instructional practices, teaching backgrounds, and characteristics. The teacher questionnaire contained two parts. The first part pertained to the teachers' background and general training. The second part pertained to specific training in teaching reading and the procedures the teacher used for *each class* containing an assessed student.

The **Teacher Questionnaire, Part I: Background and General Training** (25 questions at grade 4 and 17 at grade 8) included questions pertaining to

- gender,
- race/ethnicity,
- years of teaching experience,
- certification, degrees, major and minor fields of study,
- coursework in education,
- coursework in specific subject areas,
- amount of in-service training,
- extent of control over instructional issues, and
- availability of resources for their classroom.

This component of the questionnaire was completed by teachers whose students participated in any subject assessed in NAEP.

The **Teacher Questionnaire, Part IIA: Reading Preparation** (15 questions at grade 4 and 13 at grade 8) included questions on the teacher's exposure to various issues related to reading and teaching reading through pre- and in-service training.

The **Teacher Questionnaire, Part IIB: Reading Classroom Information** (31 questions at grades 4 and 8) included questions pertaining to

- ability level of students in the class,
- whether students were assigned to the class by ability level,
- time on task,
- homework assignments,
- frequency of instructional activities used in class,
- methods of assessing student progress in reading,
- instructional emphasis given to the reading abilities covered in the assessment, and
- use of particular resources.

2.5 DEVELOPING THE U.S. HISTORY ASSESSMENT

2.5.1 Overview

The 1994 U.S. history assessment was designed to measure a broad range of history knowledge and skills, while minimizing the burden to individual test takers. The assessment was composed of a combination of multiple-choice and constructed-response exercises, with approximately 60 percent of assessment time devoted to the latter. A broad range of materials—including primary U.S. history documents, secondary interpretations, tables, charts, graphs, artworks, photographs, and cartoons—were used as stimulus for the various questions. All test questions for the 1994 U.S. history assessment were newly developed for this survey; none were carried forward from earlier instruments.

The NAEP U.S. history framework guided the exercise development efforts. Exercises were written by ETS staff, members of the U.S. History Instrument Development Committee, and school and postsecondary teachers around the country. All assessment materials were reviewed by specialists in U.S. history education, measurement, assessment development, and bias/sensitivity. The core goals of the ETS assessment development process and the procedures used to realize these goals are outlined in the introduction to this chapter and in Section 2.3.

Sections 2.5.2 through 2.5.6 contains a detailed description of the development of the framework and exercises for the 1994 NAEP U.S. history assessment. Section 2.5.7 describes the student background and U.S. history teachers' questionnaires. Section 2.5.8 provides information about the special group study that was part of the U.S. history assessment. Additional information on the structure of the assessment booklets can be found in Chapter 4. Various committees worked on the development of the framework and exercises for the U.S. history assessment. The list of committee members who participated is included in Appendix A.

2.5.2 Development of the Assessment Framework

As in the case of the other subjects assessed in 1994, NAGB awarded a contract to the CCSSO for the development of the U.S. history assessment framework. In turn, CCSSO subcontracted with a number of other organizations—including the American History Association (AHA), the National Council for the Social Studies (NCSS), and the National Council for History Education (NCHE)—for assistance in conducting the consensus process. Specifically, these organizations were responsible for obtaining the broadest possible range of reviewers from the U.S. history and social studies education communities, so that the frameworks might represent the widest possible consensus about the history content to be covered in the NAEP survey. In addition, the CCSSO also entered into a contract with the American Institutes for Research (AIR), whose staff was charged with the preparation of the technical assessment specifications document.

The consensus process began in the summer of 1991. The process included the following general activities:

- A Steering Committee (see Appendix A) was established to provide guidance for the consensus process. The committee set guidelines for framework development, monitored the progress of the project, and offered advice. Drafts of each version of the document were sent to the Steering Committee for review.

- A Planning Committee (see Appendix A) was established to identify the objectives to be assessed in U.S. history and to prepare the framework document. The committee consisted of educators, university-level scholars, representatives from state education agencies, and noted authors. Representatives from the committee met independently and with the Steering Committee. Working in subgroups, the Planning Committee prepared the various drafts of the NAEP U.S. history framework. One subgroup assisted AIR staff in the development of assessment specifications.

- Staff from the subcontracting U.S. history and social studies education organizations (AHA, NCSS, and NCHE) conducted membership surveys and panel reviews to ensure that as many people reviewed and commented on the framework as was practically possible. In addition, to meet this same goal, public hearings and broad mail reviews on the framework were conducted, and the results of these reviews were summarized by CCSSO staff and discussed by members of the Planning Committee.

- The project staff at the CCSSO met regularly with staff from NAGB and NCES to discuss project progress.

Throughout the process, the emphasis was on involving the broadest possible range of participants. Through the efforts of CCSSO and the subcontracting organizations, literally hundreds of people participated in the framework development process.

2.5.3 Framework for the 1994 Assessment

The U.S. history framework organized and outlined the content and reasoning skills measured in the 1994 assessment. It also included achievement descriptions designed to indicate what students *should* know and be able to do in this academic area. Overall, the framework was built around the notion that NAEP should measure both concrete U.S. history knowledge and history reasoning skills. As mentioned above, special emphasis was placed on assessing students' abilities to work with a broad range of stimulus materials.

Four major U.S. history themes form the core of the history framework:

1. *Democracy* - Change and Continuity in American Democracy: Ideas, Institutions, Practices, and Controversies;

2. *Cultures* - The Gathering and Interactions of Peoples, Cultures, and Ideas;

3. *Technology* - Economic and Technological Changes and Their Relationship to Society, Ideas, and the Environment; and

4. *World Role* - The Changing Role of America in the World.

Table 2-6 shows the percentage of assessment time to be devoted to each theme, according to the assessment specifications.

Table 2-6
Percentage Distribution of Assessment Time Across U.S. History Themes as Specified in the History Framework

Grade	Democracy	Cultures	Technology	World Role
4	25%	35%	25%	15%
8	30%	30%	20%	20%
12	25%	25%	25%	25%

In addition to these themes, the framework specified the following periodization of U.S. history:

- Three Worlds and Their Meeting in the Americas (Beginnings to 1607),
- Colonization, Settlement, and Communities (1607 to 1763),
- The Revolution and the New Nation (1763 to 1815),
- Expansion and Reform (1801 to 1861),
- Crisis of the Union: Civil War and Reconstruction (1850 to 1877),
- The Development of Modern America (1865 to 1920),
- Modern America and the World Wars (1914 to 1945), and
- Contemporary America (1945 to Present).

Each period is specified as comprising a distinct proportion of the assessment. The percentages of assessment time to be devoted to each period are shown in Table 2-7.

Table 2-7
Distribution of Assessment Time Across U.S. History Periods as Specified in the Assessment Framework

Period	Grade 4	Grade 8	Grade 12
Three Worlds and Their Meeting in the Americas	20%	5%	5%
Colonization, Settlement, and Communities	15%	10%	10%
The Revolution and the New Nation	15%	20%	15%
Expansion and Reform	15%	15%	10%
Crisis of the Union: Civil War and Reconstruction	10%	20%	10%
The Development of Modern America	5%	10%	15%
Modern America and the World Wars	5%	10%	15%
Contemporary America	15%	10%	20%

A cognitive dimension measuring either U.S. history knowledge and perspective or history analysis or interpretation was specified. The percentage of assessment time to be devoted to each dimension is shown in Table 2-8.

Table 2-8
Distribution of Assessment Time Across Cognitive Dimensions as Specified in the U.S. History Assessment Framework

Grade	U.S. History Knowledge and Perspective	U.S. History Analysis and Interpretation
4	40%	60%
8	35%	65%
12	30%	70%

The assessment specifications also called for certain components of the assessment to be in-depth, or thematic, examinations of core U.S. history topics. This request was achieved through the creation of 50-minute theme blocks. Finally, the specifications indicated that no more than 50 percent of student testing time be spent on machine-scorable multiple-choice items.

2.5.4 Developing the Cognitive Items

Once a working draft of the assessment specifications was approved, the item development process began. Between March and July 1992, roughly 600 exercises were written, reviewed, and prepared for pretesting. The exercises were written by ETS U.S. history assessment development specialists, members of the NAEP U.S. History Instrument Development Committee, and school and university instructors around the country. The special thematic blocks were created by teams of ETS specialists working directly with members of the Development Committee. Writing assignment exercises were created in a format that ensured adequate coverage of the content framework.

One of the key steps in the exercise development process was the identification of appropriate stimulus materials. As mentioned earlier, the specifications called for the survey to measure the ability of students to interpret a broad range of primary and secondary U.S. history documents. In the field test, over 60 percent of the questions required students to work directly with these sorts of materials.

The assessment included constructed-response (short, regular, and extended) and multiple-choice exercises. Specific item types were used to measure given U.S. history knowledge and skills if they were judged, by ETS development specialists, to be the exercise types best suited to accurately and completely measure those skills. Both types of constructed-response questions were used to measure students' abilities to interpret documents, make simple U.S. history arguments, and show the ability to recall and use history evidence. Short constructed-response exercises were used when the U.S. history knowledge or reasoning measured could be shown in one or two sentences. Extended-answer questions—those that permitted a full page of response—were used when students were asked to make more detailed U.S. history arguments that involved marshalling evidence. In all cases, constructed-response questions were carefully scaffolded to encourage the best and most complete responses and give respondents a clear indication of what was expected of them. Multiple-choice exercises were used when a given skill or a component of U.S. history knowledge could be measured through the use of such items.

A carefully developed and proven series of steps was used to create the assessment exercises. These steps are described above, in Section 2.3.

2.5.5 Development of the Operational Forms

U.S. history field tests were conducted in February and March 1993. The field test consisted of 25-minute "survey" blocks and, at grades 8 and 12, 50-minute "thematic" blocks. Roughly 600 exercises were developed and field tested, and approximately 500 students answered every question in the field test. After field testing, responses were scored and item analyses (both classical and IRT) were conducted.

After field test analysis was completed, all items in 25-minute blocks were put in a central pool. Exercises were selected for the operational assessment and assembled into operational blocks. Items were selected for the operational assessment based on their statistical performance and the extent to which they contributed to good overall coverage of the framework. Since the 50-minute blocks were thematic, items in these blocks were not "repooled." However, some thematic block items that were not used were later included in 25-minute blocks in the main assessment.

Like the field test, the operational assessment included both 25- and 50-minute blocks. Most students were asked to respond to two 25-minute blocks; a smaller number at grades 8 and 12 responded to one 50-minute block. The grade 4 assessment included six 25-minute blocks. The instruments at grades 8 and 12 included eight 25-minute blocks and one 50-minute block. The "average" number of items in a 25-minute block at each grade is shown in Table 2-9.

Table 2-9

"Average" Number of Items in 25-Minute U.S. History Blocks

Grade	**Number of Items per Block**[5]		
	Multiple-Choice	**Short Constructed-Response**	**Extended Constructed-Response**
4	10	4	1
8	12	4	1
12	12	3	2

2.5.6 Distribution of Assessment Items

Table 2-10 lists the number of items at each grade level in the 1994 assessment. Of the total 331 questions, there are 220 unique multiple-choice items, 79 unique constructed-response questions, and 32 extended constructed-response items that make up the 1994 reading assessment. Some of these items are used at two grade levels. As a result, the sum of the items that appear at each grade (94 at grade 4, 148 at grade 8, and 156 at grade 12, for a total of 401) is greater than the total number of unique items.

Every attempt was made to achieve congruence between the pool of cognitive items chosen for the 1994 assessment and the assessment specifications. The specifications call for no more than half of student assessment time to be spent on multiple-choice questions. Table 2-11 summarizes the assessment in terms of the approximate amount of testing time allocated to different item types. While any such summary must be imprecise in an assessment in which individual items are not timed for students, to allow for calculations it was assumed that multiple-choice questions take students approximately one minute, short constructed-response questions take roughly two minutes, and extended constructed-response tasks take about five minutes.

---

[5] There is some variation between blocks; however, these numbers are accurate in most cases.

Table 2-10

Distribution of Questions for the 1994 U.S. History Assessment

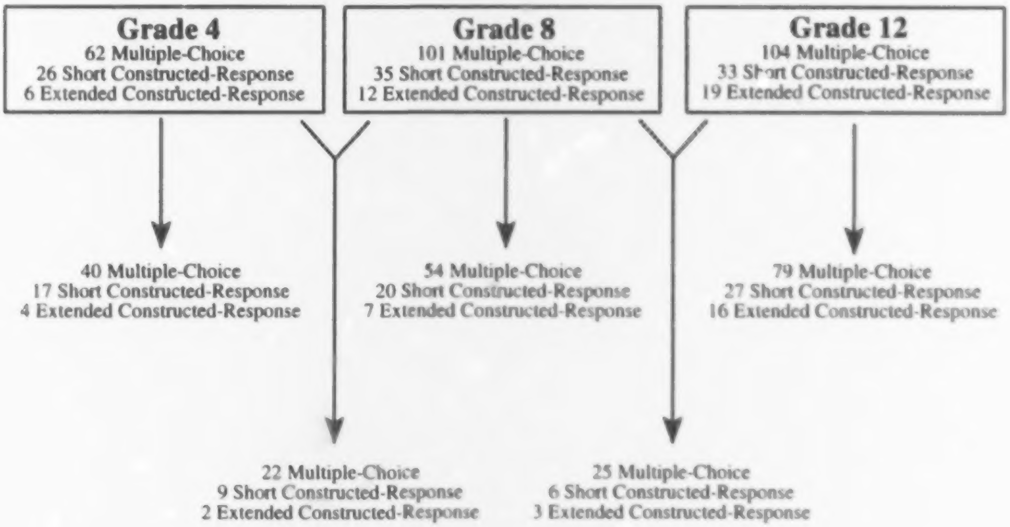

Grade 4
62 Multiple-Choice
26 Short Constructed-Response
6 Extended Constructed-Response
Grade 8
101 Multiple-Choice
35 Short Constructed-Response
12 Extended Constructed-Response
Grade 12
104 Multiple-Choice
33 Short Constructed-Response
19 Extended Constructed-Response
40 Multiple-Choice
17 Short Constructed-Response
4 Extended Constructed-Response
54 Multiple-Choice
20 Short Constructed-Response
7 Extended Constructed-Response
79 Multiple-Choice
27 Short Constructed-Response
16 Extended Constructed-Response
22 Multiple-Choice
9 Short Constructed-Response
2 Extended Constructed-Response
25 Multiple-Choice
6 Short Constructed-Response
3 Extended Constructed-Response

Table 2-11

Distribution of Assessment Time by U.S. History Item Format

Grade	Multiple-Choice	Short Constructed-Response	Extended Constructed-Response
4	43%	36%	21%
8	43%	31%	25%
12	39%	25%	36%

***Note**: Percentages may not total 100 due to rounding.*

Table 2-12 summarizes the distribution of the assessment time across the four U.S. history themes. Since these themes make up the core of the U.S. history framework, and because they will form the basis of NAEP subscores, all care has been taken to ensure the greatest possible congruence between the proportions used in the assessment and those specified in the assessment specifications. Some difficulties at grades 8 and 12 were caused by the theme blocks, which have the effect of artificially boosting coverage of a certain theme or period. Note that the numbers presented in Table 2-12 differ from Table 2-6 in that Table 2-6 shows the distribution of assessment time as specified in the U.S. history framework.

Table 2-12

Distribution of Assessment Time by U.S. History Theme

Grade	Democracy	Cultures	Technology	World Role
4	24%	32%	25%	19%
8	28%	30%	23%	19%
12	29%	23%	26%	21%

Table 2-13 summarizes the distribution of assessment time by U.S. history period. All attempts were made to conform to the specified periodization as carefully as possible. However, problems discovered during field testing made an exact compliance difficult. Note that the numbers presented in Table 2-13 differ from Table 2-7 in that Table 2-7 shows the distribution of assessment time as specified in the U.S. history framework.

Table 2-13

Distribution of Assessment Time by U.S. History Period

Period	Grade 4	Grade 8	Grade 12
NA*	13%	10%	0%
Three Worlds and Their Meeting in the Americas	15%	3%	2%
Colonization, Settlement, and Communities	13%	7%	8%
The Revolution and the New Nation	12%	20%	10%
Expansion and Reform	10%	10%	7%
Crisis of the Union: Civil War and Reconstruction	8%	13%	10%
The Development of Modern America	9%	13%	15%
Modern America and the World Wars	5%	12%	29%**
Contemporary America	15%	11%	20%

*Note: A number of questions (especially at grades 4 and 8) assess general U.S. history skills and cannot be appropriately classified in any given period. For example, one question asks students what they might learn from visiting a cemetery. The presence of these types of questions has the effect of reducing the portion of the assessment from any given period. For example, if one excluded these questions from the grade 4 calculations, each period would make up an additional two percent of the assessment, so Three Worlds and Their Meeting in the Americas would be 17 percent, Colonization, Settlement and Communities, would be 15 percent, and so forth. At grade 8, without these items, each period would make up roughly an additional two percent, as well.

**Note: This high percentage is largely a function of the 50-minute theme block that focuses on the Great Depression. If this block were excluded from calculations, questions covering this period would make up 14 percent of the assessment at this grade level.

The thematic blocks at grades 8 and 12 tended to artificially increase the number of items in a given period; this effect was most apparent in *The Development of Modern America (1865 to 1920)*, at grade 8, and in *Modern America and the World Wars (1914 to 1945)*, at grade 12. Thematic blocks tend, by their nature, to emphasize particular periods, and it is not possible simply to reduce the coverage of that period in the 25-minute blocks, given the presence of many important topics in that period. For example, the presence at grade 12 of a 50-minute block on the Great Depression does not logically imply that the assessment should not cover World Wars I and II. Further, the overlapping nature of the periods made precise classification difficult. Finally, the imperative of building the test to meet other components of the specifications necessitated some level of variance from the specified periodization.

Table 2-14 summarizes the assessment in terms of percentage of time devoted to different cognitive dimensions. It should be noted that the classification of items into different cognitive categories—conducted by both ETS staff and members of the assessment development committee—is likely an imprecise process. While every effort was made to meet the specified targets, the extensive use of constructed-response testing and primary U.S. history materials tended to increase the percentage of the assessment devoted to higher order thinking. Note that the numbers presented in Table 2-14 differ from Table 2-8 in that Table 2-8 shows the distribution of assessment time as specified in the U.S. history framework.

Table 2-14
Distribution of Assessment Time by U.S. History Cognitive Dimension

Grade	U.S. History Knowledge and Perspective	U.S. History Analysis and Interpretation
4	37%	63%
8	27%	73%
12	22%	78%

2.5.7 U.S. History Background Questionnaires

As in other NAEP subjects, questionnaires allow for the provision of a context for cognitive assessment results. The U.S. history assessment included both the general background questionnaires given to participants in all subjects and content-area specific questionnaires for both students and, at grades 4 and 8, their teachers. The development of the general background questionnaires is discussed below. It is worth noting that members of the U.S. History Instrument Development Committee were consulted on the appropriateness of all issues addressed in questionnaires that may relate to instruction and achievement. Like the cognitive items, all background questions were submitted for extensive review and field testing. Recognizing the validity problems inherent in self-reported data, particular attention was given to developing questions that were meaningful and unambiguous and that would encourage accurate reporting.

In addition to the cognitive questions, the 1994 assessment included two five-minute sets of general and U.S. history background questions designed to gather information about students' instructional experiences in history. Students in the fourth grade were given additional time because the items in the general questionnaire were read aloud to them. A one-minute questionnaire was given to students at the end of each booklet to determine students' motivation in completing the assessment and their familiarity with assessment tasks.

➪ Student U.S. History Questionnaires

Three categories of information were represented in the five-minute sections of U.S. history background questions called the **student U.S. history questionnaire** (17 questions at grade 4, 23 questions at grade 8, and 25 questions at grade 12):

> *Time Spent Studying U.S. History*: Students were asked to describe both the amount of instruction they received in U.S. history and the time spent on history homework.
>
> *Instructional Practices*: Students were asked to report their instructional experiences related to U.S. history in the classroom, including group work, special projects, and writing about history. In addition, they were asked about the instructional practices of their teachers and the extent to which the students themselves discussed U.S. history and demonstrated and practiced history reasoning skills.

➪ U.S. History Teacher Questionnaire

To supplement the information on instruction reported by students, the U.S. history or social studies teachers of the fourth and eighth graders participating in the NAEP U.S. history assessment were asked to complete a questionnaire about their instructional practices, teaching backgrounds, and characteristics. The teacher questionnaire contained two parts. The first part pertained to the teachers' background and general training. The second part pertained to specific training in teaching U.S. history and the procedures the teacher used for assessed students.

The **Teacher Questionnaire, Part I: Background and General Training** (25 questions at grade 4 and 17 at grade 8) included questions pertaining to

- gender,
- race/ethnicity,
- years of teaching experience,
- certification, degrees, major and minor fields of study,
- coursework in education,
- coursework in specific subject areas,
- amount of in-service training,
- extent of control over instructional issues, and
- availability of resources for their classroom.

This component of the questionnaire was completed by teachers whose students participated in any subject assessed in NAEP.

The **Teacher Questionnaire, Part IIIA: U.S. History Teacher Preparation** (six questions at grades 4 and 8) included questions on the teacher's preparation and comfort in teaching U.S. history.

The **Teacher Questionnaire, Part IIIB: U.S. History Classroom Information** (35 questions at grades 4 and 8) included questions pertaining to

- ability level of students in the class,
- whether students were assigned to the class by ability level,
- time on task,
- homework assignments,
- frequency of instructional activities used in class,
- methods of assessing student progress in U.S. history,
- instructional emphasis given to the content and skills covered in the assessment, and
- use of particular resources.

2.5.8 Pilot Study in Group Assessment

2.5.8.1 General Description of the Study

Increasingly, teachers in our nation's schools use group learning techniques. Cooperative learning and other heuristic models have led many educators and curriculum developers to believe that instructional settings where students help each other learn are more effective than traditional, individual-based practices alone. In addition, a number of studies have provided evidence that cooperative and group-learning experiences are positively related to higher levels of academic achievement than are individualistic or competitive instructional settings.[6] Many believe that, in addition to being an optimal means of helping students gain both knowledge and critical-thinking competencies, group learning provides students with the team-work and leadership skills necessary for success in our changing economy.[7]

While these factors suggest the importance of measuring the performance of groups of students, little work has been done assessing group work, and there is a general lack of understanding in the assessment community of the issues and challenges involved in the assessment of groups of students. For example, assessment developers have little experience in crafting exercises that are both open enough to allow for the real manifestation of group dynamics and constrained enough to meet the demands imposed by a standardized assessment environment. Nevertheless, as NAEP moves toward the future, the assessment of groups is likely to become an increasingly pressing concern. Therefore, the 1994 group assessment study in U.S. history was designed to provide initial study of the procedures and issues involved in this type of measurement.

The pilot study in group learning provided experience in the design, administration, scoring, and analysis of group assessment tasks. The study was designed to provide information on the feasibility of group assessment, and on the appropriateness of certain analytic and administrative procedures for such assessment.

---

[6] See, for example, Johnson, D. W., Maruyama, G., Johnson, R. T., Nelson, D., & Skon, L. (1981). Effects of Co-Operative, Competitive and Individualistic Goal Structures on Achievement: A Meta-Analysis. Psychological Bulletin, 89, 47-62, or Johnson, D. W., & Johnson, R. T. (1987). Cooperation and Competition. Lawrence Ehrlbaum: Hillsdale, NJ.

[7] Hill, S., & Hill, T. (1990). The Collaborative Classroom: A Guide to Co-Operative Learning. Heinemann: Portsmouth, NH.

Two group assessment activities were developed by ETS staff and were administered at grade eight in April 1994. Each activity consisted of a set of structured tasks carried out by groups of five or six students, with each group producing a series of concrete, written products. Approximately 25 groups of students participated in each of the group assessment activities. Students involved in the study had participated in the main NAEP U.S. history assessment. Schools participating in the study were chosen to represent a variety of community and school types. The range of schools was intended to allow for the measurement of the administrative problems inherent in group assessment. The study was not conducted on a statistically representative sample of schools.

For each participating school, students were selected, asked to obtain signed parental permission slips, randomly assigned to groups, and asked to work with other students to complete one of the two tasks that were administered by Westat staff. In addition to completing the group learning tasks, students completed brief questionnaires assessing their group work experience.

Groups, rather than students, were the main unit of analysis for summarizing the results of this pilot study. Two distinct types of measures were obtained for each of the groups. The first measure related to the quality of the products generated by each group. Each group product was evaluated according to criteria set forth in standardized scoring rubrics. Examples of the type of criteria used for rating purposes are quality of U.S. history thinking and analysis, history correctness and accuracy of decisions and arguments, and completeness of response to the task given.

The second set of measures related to behavioral aspects of the process used by each group in carrying out the tasks. Sessions were videotaped and only students with signed parental permission were included. Observational protocols were used to record and evaluate the nature of the interactions that occurred within each group. The process was categorized by the nature of a group's decision-making processes, the extent to which all group members share in the work, and the extent to which the group works in a cooperative and efficient fashion. Observers evaluated the process as it occurred; in addition, a process rating of the tapes was conducted after the field administration.

2.6 DEVELOPING THE GEOGRAPHY ASSESSMENT

2.6.1 Overview

The 1994 geography assessment was the first systematic assessment of the geography learning of America's youth at three grade levels. Like other NAEP assessments, the geography survey was designed to measure a broad range of knowledge and skills while minimizing the burden to individual test takers. The assessment was composed of a combination of multiple-choice and constructed-response exercises, with almost 60 percent of the assessment time devoted to the latter. In addition to written responses, students were asked to engage in "production tasks" that involved creating maps, charts, graphs, and tables. A broad range of materials—including maps, atlases, tables, charts, graphs, artworks, photographs, and cartoons—were used as stimulus for the various questions. All test questions for the 1994 geography assessment were newly developed for this survey; none were carried forward from earlier instruments.

The NAEP geography framework guided the exercise development efforts. Exercises were written by ETS staff, members of the Geography Instrument Development Committee, and school and postsecondary teachers around the country. All assessment materials were reviewed by specialists in geography education, measurement, assessment development, and bias/sensitivity. The core goals of the ETS assessment development process and the procedures used to realize these goals are outlined in the introduction to this chapter and in Section 2.3.

Sections 2.6.2 through 2.6.6 contain a detailed description of the development of the framework and exercises for the 1994 NAEP geography assessment. Section 2.6.7 describes the student background and geography teachers' questionnaires. Additional information on the structure of the assessment booklets can be found in Chapter 4. Various committees worked on the development of the framework and exercises for the geography assessment. The list of committee members who participated is included in Appendix A.

2.6.2 Development of the Assessment Framework

Like other subjects assessed in 1994, NAGB awarded a contract to the CCSSO for the development of the geography assessment framework. In turn, CCSSO subcontracted with a number of other organizations—including the National Council for Geographic Education, and NCSS—for assistance in conducting the consensus process. Specifically, these organizations were responsible for obtaining the broadest possible range of reviewers from the geography and social studies education communities, so that the frameworks might represent the widest possible consensus about the geography content to be covered in the NAEP survey. Besides these formal subcontracts, CCSSO also maintained close contact with the National Geographic Society, the American Geographical Society, and the Association of American Geographers. Representatives of these organizations sat on the consensus panel in an *ex officio* role. In addition, the CCSSO also entered into a contract with AIR, whose staff was charged with the preparation of the technical assessment specifications document.

The consensus process began in the summer of 1991. The process included the following general activities:

- A Steering Committee (see Appendix A) was established to provide guidance for the consensus process. The committee set guidelines for framework development, monitored the progress of the project, and offered advice. Drafts of each version of the document were sent to the Steering Committee for review.

- A Planning Committee (see Appendix A) was established to identify the objectives to be assessed in geography and to prepare the framework document. The members of this committee consisted of educators, university-level scholars, representatives from state education agencies, and noted authors. Representatives from the committee met independently and with the Steering Committee. Working in subgroups, the Planning Committee prepared the various drafts of the NAEP geography framework. One subgroup assisted AIR staff in the development of assessment specifications.

- Staff from the subcontracting geography and social studies education organizations conducted membership surveys and panel reviews to ensure that as many people reviewed and commented on the framework as was practically possible. In addition, to meet this same goal, public hearings and broad mail reviews on the framework were conducted, and the results of these reviews were summarized by CCSSO staff and discussed by members of the Planning Committee.

- The project staff at the CCSSO met regularly with staff from NAGB and NCES to discuss project progress.

Throughout the process, the emphasis was on involving the broadest possible range of participants. Through the efforts of CCSSO and the subcontracting organizations, over 200 people participated in the framework development process.

2.6.3 Framework for the 1994 Assessment

The geography framework organized and outlined the content and skills measured in the 1994 assessment. It also included achievement descriptions designed to indicate what students *should* know and be able to do in this academic area. Overall, the framework was built around the notion that NAEP should measure both concrete geography knowledge and spatial reasoning skills. As mentioned above, special emphasis was placed on assessing students' abilities to work with a broad range of stimulus materials, including maps and atlases, and to create maps and visual representations of their own.

Three major content areas form the core of the geography framework, and it is these content areas that yield reporting subscores. Detailed descriptions of the content categories can be found in the assessment framework. The content areas are

- Space and Place,
- Environment and Society, and
- Spatial Dynamics and Connections.

Table 2-15 shows the percentage of assessment time to be devoted to each content area, according to the assessment specifications.

Table 2-15
Percentage of Assessment Time Across Content Areas
as Specified in the Geography Framework

Grade	Space and Place	Environment and Society	Spatial Dynamics and Connections
4	40%	30%	30%
8	40%	30%	30%
12	40%	30%	30%

In addition to themes, the framework specifies a cognitive dimension for the assessment. Exercises are supposed to measure either knowing, understanding, or, applying. The level of applying is designed to refer to a range of higher order skills. The percentages of the assessment time to be devoted to each cognitive dimension are shown in Table 2-16.

Table 2-16
Percentage of Assessment Time Across Cognitive Dimensions as Specified in the Geography Framework

Grade	Knowing	Understanding	Applying
4	45%	30%	25%
8	40%	30%	30%
12	30%	30%	40%

The framework also placed a major emphasis on the use in the assessment of a variety of maps, charts, graphs, atlases, tables, and photographs. These materials, which serve as stimuli for questions and tasks, were used to measure students' ability to interpret and analyze the spatial materials that make up the core of geography thinking.

The framework further instructed that some students work in-depth on a single geography theme or topic. This specification was met through the development of 50-minute theme blocks. Finally, the specifications indicated that no more than 50 percent of student testing time be spent on machine-scorable multiple-choice items. This specification was adhered to in the construction of the assessment.

2.6.4 Developing the Cognitive Items

Once a working draft of the assessment specifications had been approved, the item development process began. Between March and July 1992, almost 500 exercises were written, reviewed, and prepared for field testing. The exercises were written by ETS assessment development specialists, members of the NAEP Geography Instrument Development Committee, and school and university instructors around the country. The special thematic blocks were created by teams of ETS specialists working directly with members of the Development Committee. Exercise writing assignments were created in a format that ensured adequate coverage of the content framework.

One of the key steps in the exercise development process was the identification of appropriate stimulus materials. As mentioned earlier, the specifications called for the survey to measure the ability of students to interpret a broad range of maps, photographs, graphics, tables, charts, and cartoons. In addition, the Geography Instrument Development Committee members worked with ETS Development staff to select an appropriate school atlas for the atlas components of the assessment. In the field test, over 80 percent of the questions required students to work directly with a visual stimulus.

The assessment included constructed-response (short, regular, and extended) and multiple-choice exercises. Specific item-types were used to measure given knowledge and skills if they were judged, by ETS development specialists, to be the exercise-types best suited to accurately and completely measure those skills. Both types of constructed-response questions were used to measure students' abilities to interpret maps and charts, make spatial arguments, and show the ability to recall and use knowledge. Short constructed-response exercises were used when the geography knowledge or reasoning measured could be shown in one or two sentences. Extended-answer questions—those that permitted a full page of response—were used when students were asked to make more detailed geography arguments that involved marshalling evidence. "Production tasks" were used when the question was intended to use students' abilities to create their own maps, charts, or tables. Finally, a number of constructed-response exercises were designed to assess student abilities to apply geography reasoning to real-life problems. In all cases, constructed-response questions were carefully scaffolded encourage the best and most complete possible responses and to give respondents a clear indication of what was expected of them. Multiple-choice exercises were used when a given skill or a component of knowledge could be measured through the use of such items.

A carefully developed and proven series of steps was used to create the assessment exercises. These steps are described above, in Section 2.3.

2.6.5 Development of the Operational Forms

Geography field tests were conducted in February and March 1993. The field test consisted of 25-minute survey blocks and, at grades 8 and 12, 50-minute thematic blocks. Roughly 460 exercises were developed and field tested, and approximately 500 students answered every question in the field test. After field testing, responses were scored and item analyses (both classical and IRT) were conducted.

After field test analysis was completed, all items in 25-minute blocks were put in a central pool. Exercises were selected for the operational assessment, and these exercises were assembled into operational blocks. Items were selected for the operational assessment based on their statistical performance and the extent to which they contributed to good overall coverage of the framework. Since the 50-minute blocks were thematic, items in these blocks were not repooled. However, some thematic block items that were not used were later included in 25-minute blocks in the main assessment.

Like the field test, the operational assessment included both 25- and 50-minute blocks. Most students were asked to respond to two 25-minute blocks; a smaller number at grades 8 and 12 responded to one 50-minute block. The grade 4 assessment included six 25-minute blocks. The instruments at grades 8 and 12 included six 25-minute blocks and one 50-minute block. The "average" number of items in a 25-minute block at each grade is shown in Table 2-17.

Table 2-17

"Average" Number of Items in 25-Minute Geography Blocks

Grade	Number of Items per Block[8]		
	Multiple-Choice	**Short Constructed-Response**	**Extended Constructed-Response**
4	10	4	1
8	12	4	1
12	12	3	2

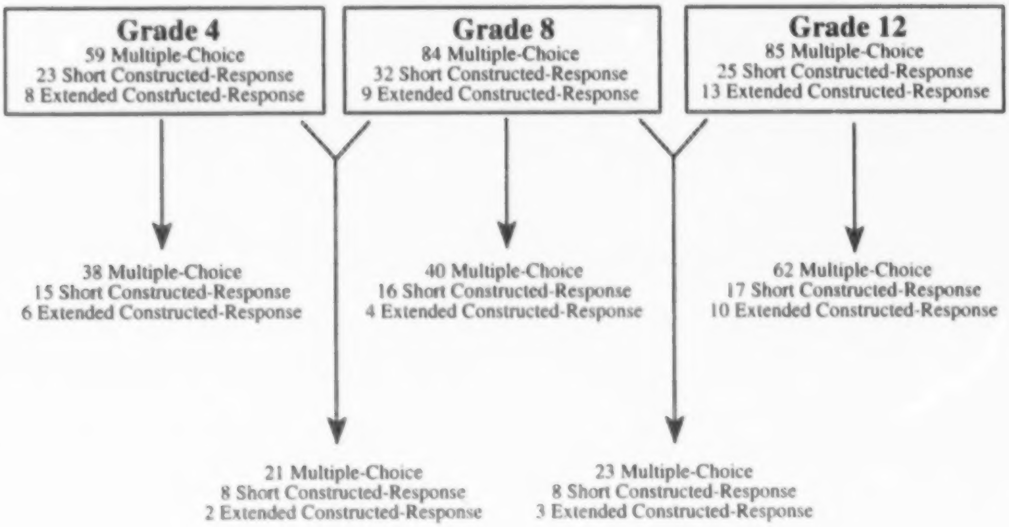
---

[8] There is some variation between blocks; however, these numbers are accurate in most cases.

2.6.6 Distribution of Assessment Items

Table 2-18 lists the total number of assessment items at each grade in the 1994 assessment. Of the total 273 questions, there are 184 unique multiple-choice items, 64 short constructed-response items, and 25 extended constructed-response items that make up the assessment. Some of these items are used at two grades. As a result, the sum of the items that appear at each grade (90 at grade 4, 125 at grade 8, and 123 at grade 12, for a total of 338) is greater than the total number of unique items.

Table 2-18
Distribution of Items for the 1994 Geography Assessment


Grade 4
59 Multiple-Choice
23 Short Constructed-Response
8 Extended Constructed-Response
Grade 8
84 Multiple-Choice
32 Short Constructed-Response
9 Extended Constructed-Response
Grade 12
85 Multiple-Choice
25 Short Constructed-Response
13 Extended Constructed-Response
38 Multiple-Choice
15 Short Constructed-Response
6 Extended Constructed-Response
40 Multiple-Choice
16 Short Constructed-Response
4 Extended Constructed-Response
62 Multiple-Choice
17 Short Constructed-Response
10 Extended Constructed-Response
21 Multiple-Choice
8 Short Constructed-Response
2 Extended Constructed-Response
23 Multiple-Choice
8 Short Constructed-Response
3 Extended Constructed-Response

Every attempt was made to achieve congruence between the pool of cognitive items chosen for the 1994 assessment and the assessment specifications. The specifications call for no more than half of student assessment time to be spent on multiple-choice questions. Table 2-19 summarizes the assessment in terms of the approximate amount of testing time allocated to different item types. While any such summary must be imprecise in an assessment in which individual items are not timed for students, to allow for calculations it was assumed that multiple-choice questions take students approximately one minute, short constructed-response questions take roughly two minutes, and extended constructed-response tasks take about five minutes. Reviewers should be aware that the actual amount of time students spend on questions will likely vary significantly within a given category.

Table 2-19

Distribution of Assessment Time by Geography Item Format

Grade	Multiple-Choice	Short Constructed-Response	Extended Constructed-Response
4	41%	32%	28%
8	44%	33%	23%
12	43%	25%	33%

Table 2-20 summarizes the distribution of the assessment time across the three content areas. Since these content categories make up the core of the geography framework, and because they will form the basis of NAEP subscores, all care has been taken to ensure the greatest possible congruence between the proportions used in the assessment and those specified in the assessment specifications. Note that the numbers presented in Table 2-20 differ from Table 2-15 in that Table 2-15 shows the distribution of assessment time as specified in the geography framework.

Table 2-20

Distribution of Assessment Time by Geography Content Area

Grade	Space and Place	Environment and Society	Spatial Dynamics and Connections
4	42%	28%	31%
8	39%	30%	32%
12	42%	30%	29%

Table 2-21 summarizes the assessment in terms of percentage of time devoted to different cognitive dimensions. It should be noted that the classification of items into different cognitive categories—conducted by both ETS staff and members of the assessment development committee—is likely an imprecise process. While every effort was made to meet the specified targets, the extensive use of constructed-response testing and varied stimulus materials and response formats tended to increase the percentage of the assessment devoted to higher order thinking. For example, while questions in the *knowing* category make up 47 percent of the questions at grade 4, they will consume only approximately 37 percent of the assessment time. Note that the numbers presented in Table 2-21 differ from Table 2-16 in that Table 2-16 shows the distribution of assessment time as specified in the geography framework.

Table 2-21

Distribution of Assessment Time and Item Pool by Geography Cognitive Dimension

Grade	Knowing		Understanding		Applying	
	% of Items	% of Time	% of Items	% of Time	% of Items	% of Time
4	47%	37%	37%	37%	17%	26%
8	40%	29%	39%	42%	20%	28%
12	40%	26%	39%	33%	21%	42%

***Note:** Percentages may not total 100 due to rounding.*

2.6.7 Geography Background Questionnaires

As in other NAEP subjects, questionnaires allow for the provision of a context for cognitive assessment results. The geography assessment included both the general background questionnaires given to participants in all subjects and content-area specific questionnaires for both students and, at grades 4 and 8, their teachers. The development of the general background questionnaires is discussed below. It is worth noting that members of the Geography Instrument Development Committee were consulted on the appropriateness of all issues addressed in questionnaires that may relate to instruction and achievement. Like the cognitive items, all background questions were submitted for extensive review and field testing. Recognizing the validity problems inherent in self-reported data, particular attention was given to developing questions that were meaningful and unambiguous and that would encourage accurate reporting.

In addition to the cognitive questions, the 1994 assessment included two five-minute sets of general and geography background questions designed to gather information about students' instructional experiences in geography. A one-minute questionnaire was given to students at the end of each booklet to determine students' motivation in completing the assessment and their familiarity with assessment tasks.

➪ Student Geography Questionnaires

Three categories of information were represented in the five-minute sections of reading background questions called the **student geography questionnaire** (22 questions at grade 4, 25 questions at grade 8, and 29 questions at grade 12):

Time Spent Studying Geography: Students were asked to describe both the amount of instruction they received in geography and the time spent on geography homework.

Instructional Practices: Students were asked to report their instructional experiences related to geography in the classroom, including group work, special projects, and writing about geography. In addition, they were asked about the instructional practices of their teachers and the extent to which the students themselves discussed geography and demonstrated and practiced geography reasoning skills.

Attitudes Towards Geography: Students' enjoyment of geography was measured through questions about their attitudes about this subject.

➪ Geography Teacher Questionnaires

To supplement the information on instruction reported by students, the geography or social studies teachers[9] of the fourth and eighth graders participating in the NAEP geography assessment were asked to complete a questionnaire about their instructional practices, teaching backgrounds, and characteristics. The teacher questionnaire contained two parts. The first part pertained to the teachers' background and general training. The second part pertained to specific training in teaching geography and the procedures the teacher used for assessed students.

The **Teacher Questionnaire, Part I: Background and General Training** (25 questions at grade 4 and 17 at grade 8) included questions pertaining to

- gender,
- race/ethnicity,
- years of teaching experience,
- certification, degrees, major and minor fields of study,
- coursework in education,
- coursework in specific subject areas,
- amount of in-service training,
- extent of control over instructional issues, and
- availability of resources for their classroom.

This component of the questionnaire was completed by teachers whose students participated in any subject assessed in NAEP.

The **Teacher Questionnaire, Part IVA: Geography Teacher Preparation** (nine questions at grade 4 and eight questions at grade 8) included questions on the teacher's preparation and comfort in teaching geography.

---

[9] Note that if there was no geography teacher, the appropriate social studies or U.S. history teacher completed the questionnaire.

The **Teacher Questionnaire, Part IVB: Geography Classroom Information** (28 questions at grade 4 and 29 questions at grade 8) included questions pertaining to

- ability level of students in the class,
- whether students were assigned to the class by ability level,
- time on task,
- homework assignments,
- frequency of instructional activities used in class,
- methods of assessing student progress in geography,
- instructional emphasis given to the geography skills and content covered in the assessment, and
- use of particular resources.

2.7 BACKGROUND QUESTIONNAIRES

As part of NAEP, a series of questionnaires was administered to students, teachers, and school administrators. Similar to the development of the cognitive items, the development of the policy framework and questionnaire items was a consensual process that involved staff work, field testing, and review by external advisory groups.

Before the 1992 assessment, a Background Questionnaire Committee was appointed to oversee the development of a policy framework for background questionnaires. The policy framework, which was again used as part of the development efforts for the 1994 assessment, focused on five educational areas: instructional content, instructional practices and experiences, teacher characteristics, school conditions and contexts, and conditions outside of school. The panel was also interested in capitalizing on the unique qualities of NAEP and not duplicating other surveys.

For the 1994 assessment, a new Background Questionnaire Committee oversaw the development of the various questionnaires. The panel recommended that study be focused on areas that addressed the relationship between student achievement and instructional practices. The policy issues, items, and field test results were reviewed by a group of external consultants who identified specific items to be included in the final questionnaires. In addition, members of the Instrument Development Committee in all three subjects reviewed the questionnaires. The questionnaires underwent ETS reviews designed to ensure quality and fairness. All questions were systematically field tested.

2.7.1 Student Questionnaires

In addition to the cognitive items, the 1994 assessment included two five-minute sets of general and subject-specific background items (the same two sets of items that were described in Section 2.6.7) designed to gather contextual information about students, their educational experiences, and their perceptions of the subjects assessed. Students in the fourth grade were given additional time because the items in the general questionnaire were read aloud for them. Students were also given a one-minute set of background questions designed to measure their motivation on the assessment. In many cases the questions used were continued from prior assessments in order to measure change over time.

The **student demographics (common core) questionnaire** included questions about race/ethnicity, language spoken in the home, mother's and father's level of education, reading materials in the home, television watching, homework, and which parents live at home. This questionnaire was included in the first section of every booklet.

The **student subject-specific questionnaires** covered issues such as time spent studying a specific subject, coursework, the nature of students' instruction, and students' enjoyment of and confidence in their abilities in subject areas, and their perceptions of the usefulness of the disciplines to their present and future lives. This questionnaire was included in the fourth section of every booklet.

The **motivation questionnaire** asked the students questions about their perceptions of the difficulty of the assessment, how well they think they did on the assessment, and their motivation to do well on the assessment. This questionnaire was included in the last section of every booklet.

2.7.2 Teacher, School, and IEP/LEP Student Questionnaires

To supplement the information on instruction reported by students, questionnaires were completed by the reading, U.S. history, and geography teachers of fourth- and eighth-grade students assessed in those subject areas. The teacher questionnaires contained two parts. The first part pertained to the teachers' background and general training. The second part pertained to specific training in teaching the subjects being assessed, and the procedures the teacher used while giving instruction to assessed students.

Because the sampling for the teacher questionnaires was based on participating students, the responses to a particular teacher questionnaire do not necessarily represent all teachers of that subject area at that grade level in the nation. Rather, they are teachers of the representative sample of students assessed. It is important to note that in this report, as in all NAEP reports, the student is always the unit of analysis, even when information from the teacher or school questionnaire is being reported. Using the student as the unit of analysis makes it possible to describe the instruction received by representative samples of students. Although this approach may provide a different perspective from other studies simply reporting information about teachers or schools, it is consistent with NAEP's goal of providing information about the educational context and performance of students.

The **Teacher Questionnaire, Part I: Background and General Training** included questions pertaining to

- gender,
- race/ethnicity,
- years of teaching experience,
- certification, degrees, major and minor fields of study,
- coursework in education,
- coursework in specific subject areas,
- amount of in-service training,
- extent of control over instructional issues, and
- availability of resources for their classroom.

The **Teacher Questionnaire, Part II: Subject-Specific Information** included questions pertaining to

- ability level of students in the class,
- whether students were assigned to the class by ability level,
- time on task,
- homework assignments,
- frequency of instructional activities used in class,
- methods of assessing student progress in reading,
- instructional emphasis given to the reading abilities covered in the assessment, and
- use of particular resources.

The **School Characteristics and Policies Questionnaire** was given to the principal or administrator of each school that participated in NAEP. It included questions pertaining to

- background and characteristics of school principals,
- length of school day and year,
- school enrollment,
- absenteeism,
- drop-out rates,
- policies about tracking,
- curriculum,
- testing practices and use,
- special priorities and school-wide programs,
- availability of resources, special services, and community services,
- policies for parental involvement, and
- school-wide problems.

The **IEP/LEP Student Questionnaire** was completed by the teachers of those students who were selected to participate in the assessment sample who had Individualized Education Plans (IEP) or were classified as Limited English Proficient (LEP). The questionnaire was completed for all IEP or LEP students, whether or not they actually participated in the assessment. This questionnaire asked about the nature of the student's disability and the special programs in which the student participated.

Schools were permitted to exclude certain students from the assessment. The same exclusion criteria and rules used in the national assessment were also applied to the Trial State Assessment. Although the intent was to assess all sampled students, students who were identified by school staff as not capable of participating meaningfully were excluded. The NAEP guidelines for exclusion were intended to assure uniformity of exclusion criteria from school to school as well as from state to state.

Chapter 3

SAMPLE DESIGN[1]

Leslie Wallace and Keith F. Rust
Westat, Inc.

3.1 INTRODUCTION

The samples for the 1994 NAEP assessment were selected using a complex multistage sample design involving the sampling of students from selected schools within 94 selected geographic areas, called primary sampling units (PSUs), across the United States.

The sample design had four stages of selection:

1. selection of geographic PSUs (counties or groups of counties),
2. selection of schools within PSUs,
3. assignment of session types to schools, and
4. selection of students for session types within schools.

The samples were drawn for the three different age classes[2], and for each age class the samples were of two distinct types. The first type consisted of the cross-sectional or "main" samples, while the second type consisted of the long-term trend samples. The populations surveyed with each of these sample types are defined in Table 1-1. Separate samples of schools were required for the long-term trend samples and main samples, because of various differences in the calendar period for test administration, the format of the administration, and, in the case of age class 17, the grade and age definition of the population of interest. (See the description of Table 1-1 in Chapter 1.)

In addition to representing the respective populations as a whole, for the main samples there was oversampling of nonpublic schools, and of public schools with moderate or high enrollment of Black or Hispanic students (see Section 3.3). This oversampling was undertaken to increase the sample sizes of nonpublic school students and minority students, so as to increase the reliability of estimates for these groups of students.

The overall assessment period fell into three time periods—fall, winter, and spring. Not all assessment components were conducted in each time period. Table 3-1 shows the relationship between the various sample components and the assessment periods. The sizes of the PSU and school samples and the procedures for their selection were determined by the assessment period, as well as by the population to be surveyed and the method of administration in each case.

---

[1] Leslie Wallace and Keith F. Rust are responsible for the design and implementation of the sampling process for NAEP.

[2] The term "age class" is used in this report when it is appropriate to discuss one of the three student cohorts in a general way (not necessarily in reference to a specific sample). For the 1994 assessment, age class 9 refers to age 9 and age 9/grade 4 students, age class 13 refers to age 13 and age 13/grade 8 students, and age class 17 includes the age 17, age 17/grade 11, and age 17/grade 12 students.

Table 3-1

Assessment Type by Age Class and Assessment Period

Age Class	**Fall** **10/11/93 - 12/17/93**	**Winter** **1/3/94 - 3/11/94**	**Spring** **3/14/94 - 5/13/94**
9	—	Main assessment Long-term trend assessment	—
13	Long-term trend assessment	Main assessment	—
17	—	Main assessment	Long-term trend assessment

The age class 9 and age class 13 long-term trend samples used the same school and student eligibility requirements as the respective main samples. Nevertheless, special trend samples were required because:

1. The conditions for administration of the assessment varied considerably between the main sample and long-term trend sample sessions.

2. The need in the long-term trend samples for four distinct session types for age class 9 and 13 and three for age class 17, together with the need for two distinct session types for the main samples, made it infeasible to conduct both main sample sessions and long-term trend sessions in a given school. The session types were spiral and tape booklets 91-93 for ages 9 and 13, spiral and tape booklets 84-85 for age 17 long-term trend, and reading and U.S. history/geography for the main samples.

3. For age class 13, the main samples were conducted at an inappropriate time of the year for trend purposes, so that a distinct sample of schools was needed to undertake the long-term trend assessments for age class 13.

A separate sample of schools was required for the long-term trend sessions and the main sessions for age class 17 primarily because the definitions for student eligibility, based on age and grade, differed substantially between the two samples, even though the same population of schools was surveyed in each case. Conditions of administration also varied somewhat, and there were up to four distinct trend session types and two distinct main sample session types per age class. Thus, it was not feasible to conduct main sample and long-term trend sample sessions within a single school.

This chapter gives details of the sample selection procedure, and information on the results of the sampling process. Further details are given in the report *1994 National Assessment of Educational Progress Sampling and Weighting Procedures, Final Report* (Wallace & Rust, 1996).

3.2 PRIMARY SAMPLING UNITS

In the first stage of sampling, the United States (the 50 states and the District of Columbia) was divided into geographic primary sampling units (PSUs). The PSUs are different from those used in 1992 and earlier years, and incorporate 1990 Census information. With a few exceptions, each PSU met a minimum size requirement (a 1990 Census population of at least 60,000 in the Northeast and Southeast and 45,000 in the Central and West regions) and comprised either a consolidated metropolitan statistical area (CMSA), a metropolitan statistical area (MSA), a single county, or (more likely in the case of nonMSA PSUs) a group of contiguous counties. In the case of New England MSAs, which are not formed from whole counties, the corresponding New England County Metropolitan Areas, which are defined in terms of whole counties, were designated as the PSUs. Each PSU was contained entirely within one of the four NAEP regions defined in Table 3-2. These NAEP regions were used to stratify the sample of PSUs, ensuring that each region was adequately represented in the various assessment samples.

Table 3-2

Geographic Regions Used for Stratification

Northeast	Southeast	Central	West
Connecticut	Alabama	Illinois	Alaska
Delaware	Arkansas	Indiana	Arizona
District of Columbia	Florida	Iowa	California
Maine	Georgia	Kansas	Colorado
Maryland	Kentucky	Michigan	Hawaii
Massachusetts	Louisiana	Minnesota	Idaho
New Hampshire	Mississippi	Missouri	Montana
New Jersey	North Carolina	Nebraska	Nevada
New York	South Carolina	North Dakota	New Mexico
Pennsylvania	Tennessee	Ohio	Oklahoma
Rhode Island	Virginia*	South Dakota	Oregon
Vermont	West Virginia	Wisconsin	Texas
Virginia*			Utah
			Washington
			Wyoming

Note: *That part of Virginia that is part of the Washington, DC-MD-VA metropolitan statistical area is included in the Northeast region; the remainder of the state is included in the Southeast region.*

In a few cases an MSA crossed region boundaries. Such MSAs were split into two or more PSUs as necessary (e.g., the Cincinnati OH-KY-IN MSA was split into the Cincinnati OH-IN PSU in the Central region and the Cincinnati KY PSU in the Southeast). Ninety-four PSUs were selected for the main samples and 52 PSUs were selected for the long-term trend samples, as described below.

For the main samples, the 22 largest PSUs were included with certainty. The inclusion of these PSUs in the sample with certainty provided an approximately optimum, cost-efficient sample of schools and students when samples were drawn within them at the required national sampling rate. The 22 largest PSUs by region were:

Northeast:
Baltimore, MD MSA
Boston-Lawrence-Salem-Lowell-Brockton, MA NECMA
New York-Northern New Jersey-Long Island, NY-NJ CMSA
Philadelphia-Wilmington-Trenton, PA-NJ-DE-MD CMSA
Pittsburgh-Beaver Valley, PA CMSA
Washington, DC-MD-VA MSA

Southeast:
Atlanta, GA MSA
Miami-Fort Lauderdale, FL CMSA
Tampa-St. Petersburg-Clearwater, FL MSA

Central:
Chicago-Gary-Lake County, IL-IN-WI CMSA
Cleveland-Akron, MI CMSA
Detroit-Ann Arbor, MI CMSA
Minneapolis-St. Paul, MN-WI MSA
St. Louis, MO-IL MSA

West:
Dallas-Fort Worth, TX CMSA
Denver-Boulder, CO CMSA
Houston-Galveston-Brazoria, TX CMSA
Los Angeles-Anaheim-Riverside, CA CMSA
Phoenix, AZ MSA
San Diego, CA MSA
San Francisco-Oakland-San Jose, CA CMSA
Seattle-Tacoma, WA CMSA

The remaining smaller PSUs were not guaranteed to be selected for the sample. These were grouped into a number of noncertainty strata (so called because the PSUs in these strata were not included in the sample with certainty), and one PSU was selected from each stratum. The PSUs were classified into four regions, each containing about one-fourth of the U.S. population. These regions were defined primarily by state (Table 3-2). In each region, noncertainty PSUs were classified as MSA (metropolitan) or nonMSA (nonmetropolitan). The resulting major strata are shown in Table 3-3.

Table 3-3
The Sampling Major Strata
and the Number of Noncertainty Strata in Each

Region	Number of Strata for MSA PSUs	Number of Strata for NonMSA PSUs	Total Strata
Northeast	6	4	**10**
Southeast	12	12	**24**
Central	8	12	**20**
West	10	8	**18**
Total	**36**	**36**	**72**

Within each major stratum, further stratification was achieved by ordering the noncertainty PSUs according to several additional socioeconomic characteristics, yielding 72 strata. The number of such strata formed within each major stratum is shown in Table 3-3. The strata were defined so that the aggregate of the measures of size of the PSUs in a stratum was approximately equal for each stratum. The size measure used was the population from the 1990 Census. The characteristics used to define strata were the percent minority population, the percentage change in total population since 1980, the per capita income, the percent of persons age 25 or over with college degrees, the percent of persons age 25 or over who completed high school, and the civilian unemployment rate. Up to four of these characteristics were used in any one major stratum. For each major stratum, the characteristics used were chosen by modeling PSU-level mean reading proficiency scores for 1988, 1990, and 1992. The characteristics chosen were the best predictors of PSU-level mean reading proficiency scores in these models. One PSU was selected with probability proportional to size from each of the 72 noncertainty strata. That is, within each stratum, a PSU's probability of being selected was proportional to its population.

The final main sample of 94 PSUs was drawn from a population of about 1,000 PSUs. Primarily because of the use of MSAs as PSUs, PSUs varied considerably as to their probability of selection, since they varied greatly in size. The 36 selected noncertainty MSA PSUs had probabilities of selection ranging from 0.023 to 0.580, while the 36 selected nonMSA PSUs had probabilities ranging from 0.029 to 0.108. Parts of 41 states were included in the main sample PSUs.

Samples of 94 PSUs each were also drawn for the 1996, 1998, 2000, and 2002 assessments. They were drawn to minimize overlap of the PSUs from one assessment to the next, except that certainty PSUs were retained in each assessment year, and some of the larger noncertainty PSUs are in the sample for more than one of these assessment years.

For the long-term trend samples, 52 PSUs were selected. The long-term trend samples were much smaller than the main samples and used separate field staff. Fewer PSUs were used for the long-term trend samples to avoid having the sample spread too thinly across PSUs. The long-term trend PSUs were drawn for 1994, 1996, 1998, 2000, and 2002 to minimize overlap of the trend PSUs from one assessment to the next, and to minimize overlap between trend and main within the same assessment.

The 10 largest main sample certainty PSUs were also included with certainty in the long-term trend samples. Six additional PSUs were selected systematically and with probability proportional to the 1990 population from the 12 remaining main sample certainties. Finally, 36 PSUs were selected from the 72 noncertainty strata so that the overall procedure was equivalent to systematic sampling with probabilities proportional to the 1990 population. Note that the noncertainty long-term trend PSUs are not a subsample of the noncertainty main sample PSUs, in order to minimize the burden on a given school district in any one year.

3.3 SELECTION OF SCHOOLS

In the second stage of sampling, the public schools (including Bureau of Indian Affairs schools and Department of Defense schools) and nonpublic schools (including Catholic schools) were listed according to the grade ranges associated with the three age classes. Table 3-4 shows the numbers of schools included in the various frame components. The population of eligible public schools for each age class was the same for long-term trend and main samples in each case, except that the schools were restricted to the selected PSUs. Main sample schools were selected from the 94 main sample PSUs and long-term trend schools were selected from the 52 long-term trend PSUs. This was a change from past years, when long-term trend samples were selected from a subset of the main sample PSUs. The different administrative procedures required for the main and long-term trend assessments, coupled with the smaller sample sizes for the long-term trend assessments made it desirable to conduct the long-term trend assessments in a smaller sample of PSUs using a field staff independent of that used in the main assessments.

Table 3-4

Grade Definition of School Eligibility for Frame Inclusion and Frame Sizes, Main and Long-Term Trend Samples

	Frame Included	Number of Public Schools on Frame*		Number of Nonpublic Schools on Frame**	
Age Class	Schools with Any Grade in This Range	Long-Term Trend	Main	Long-Term Trend	Main
9	2 - 5	16,600	21,077	7,150	11,599
13	6 - 9	15,479	19,312	7,748	12,244
17	9 - 12	4,043	5,178	2,377	5,406
Total	2 - 12	23,541	29,839	8,254	13,016

Note: Public, Bureau of Indian Affairs, and Department of Defense Schools

***Note: Catholic and other nonpublic schools*

Any school having one or more of the eligible grades, and located within an appropriate PSU, was included on the frame of schools (the list of schools from which the samples of schools were drawn) for a given sample. For each age class, only a fraction of one percent of age- eligible students was enrolled in ineligible schools. An independent sample of schools was selected for each of the age classes, separately for main and long-term trend. Thus some schools were selected for assessment of two age classes, and a few were selected for all three. For all three age classes, a sample of schools was first drawn for the long-term trend assessments. These schools were then excluded from the frame when the samples of schools were drawn for the main assessments. Adjustments were made to the sampling weights to reflect the appropriate probabilities of selection to yield unbiased estimates for both long-term trend and main samples.

The lists of schools were obtained from several sources. Regular public, Bureau of Indian Affairs, Department of Defense, Catholic, and other nonpublic schools were obtained from the 1992 list of schools maintained by Quality Education Data, Inc. (QED).

For the main samples, supplementary lists of nonpublic schools were obtained from three sources and added to the QED list of nonpublic schools. This supplementation was undertaken because previous studies have revealed that the QED list of nonpublic schools is somewhat deficient in its coverage of nonCatholic nonpublic schools (Westat, Inc., 1984, Chapter 4). Although the percentage of students in schools not covered by QED is small as a percentage of the total student population (believed to be less than one percent), we believed that it was important to attempt to reduce this noncoverage for the main samples, since separate estimates were to be produced for nonpublic school students, based on oversampling of nonpublic schools.

The first supplementary nonpublic school listing source used was the Private School Survey (PSS) developed for the National Center for Education Statistics' 1988 School and Staffing Survey. This list was restricted to a sample of counties selected for the survey. Certain of these counties, generally large in population, were also included, independently by chance, in the NAEP sample PSUs. The schools from such counties were added to the NAEP frame after steps were taken to eliminate duplicates with the QED list of nonpublic schools. The second and third sources were lists generated clerically from the yellow pages of telephone directories with metropolitan areas included in both the 1992 and 1994 NAEP PSU samples, and from metropolitan areas included only in the 1994 NAEP PSU samples. These lists were matched against each other and the other nonpublic school sources to eliminate duplicates.

The process of nonpublic school list supplementation resulted in the addition of 2,896 schools to the frame: 601 obtained from the School and Staffing Survey list, 555 derived from telephone directories in 1992 and 1,740 derived from telephone directories in 1994. The nature of these listing sources meant that little was known about these schools, in particular the grade span. This meant that a large proportion of these schools were in fact out of scope for a given age class school frame. These 2,896 schools were included in the sampling process, and 220 were selected. Of the selected schools, 164 were found to be out of scope, either by preliminary screening via telephone or by the Westat field supervisor who contacted the school regarding participation in NAEP. Thus it appears that only about 26 percent of the supplementary nonpublic schools were in scope for the sample for which they were selected.

For each sample, schools were selected (without replacement) across all PSUs with probabilities proportional to assigned measures of size. In those PSUs included in both main and trend samples, the probability of selection for long-term trend for any school in a given age class was capped at 0.5, to ensure that adequate schools remained to be selected for the main sample. Equal measures of size were assigned to schools containing estimates of age-eligible students ranging from 20 to 60 for each age class. The measure of size used for each school was the estimated number of age eligible students in the school, since for each age class the large majority of students selected were assigned to sessions for which only students of the appropriate age were eligible. In most schools having the modal grade, some additional students were selected who were in the modal grade but not age-eligible, so that the maximum sample size of students within a school was about 80 grade- and age-eligible students. Schools with more than 60 age-eligible students were selected with probabilities proportional to the measure of size. Schools with fewer than 20 estimated age-eligibles were assigned somewhat lower measures of size, and thus lower probabilities of selection, since assessment in these schools involved substantially higher per-student administrative costs. For the main samples, equal measures of size were assigned to schools containing estimates of age- plus grade-eligible students ranging from 20 to 150 (for age class 9) or 20 to 200 (for age class 13 and age class 17). Schools larger than the indicated maximum size were selected with probabilities proportional to the measure of size. Schools with fewer than 20 estimated age- plus grade-eligibles were assigned somewhat lower measures of size, and thus lower probabilities of selection.

108 A

For the main samples, each public school designated as high-minority (with over 10 percent Black and/or Hispanic enrollment for age classes 9 and 13, or over 15 percent Black and/or Hispanic enrollment for age class 17) was given double the probability of selection of a public school, not designated high-minority, of similar size in the same PSU. Such high-minority schools were oversampled in order to enlarge the sample of Black and Hispanic students, thereby enhancing the reliability of estimates for these groups. For a given overall size of sample, this procedure reduces somewhat the reliability of estimates for all students as a whole and for those not Black or Hispanic. Each nonpublic school was given triple the probability of selection of a public school not designated high-minority of similar size in the same PSU. These greater probabilities of selection were used to ensure adequate samples of nonpublic school students in order to allow the derivation of reliable estimates for such students. No subgroups (high minority schools or nonpublic schools) were oversampled in the long-term trend samples.

The total number of schools selected for each age class was such that the predesignated student sample sizes would be achieved by selecting all eligible students in a selected school, up to the maximum sizes indicated above, allowing for losses due to nonparticipation of selected schools and students and the exclusion of students from the assessment. This design, with the important exceptions described above, had the goal of yielding a sample of students in a given age or grade with approximately uniform probabilities of selection. The efforts to oversample nonpublic school students and minority students and the practical constraints on the sample size within each school resulted in some substantial violations of this general goal. The distributions of selection probabilities of the selected students, as reflected in their sampling weights, are mentioned in Chapter 10.

The QED files do not contain schools that opened between 1992 and the assessment dates. Therefore, special procedures were implemented to be sure that the NAEP assessment represented students in new public schools. Small school districts, which generally contained only one eligible school for a given age class, were handled differently from large school districts, which generally contained more than one eligible school for a given age class. In small school districts, the schools selected for a given age class were thought to contain all students in the district that were eligible for the assessment. Districts containing these schools were asked if other schools with the appropriate grades for the assessment existed, and if so, they were automatically included in the assessment. For large school districts, a district-level frame was constructed from the schools on the QED file that were eligible for one of the national assessments. Then districts were sampled systematically with probabilities proportional to a measure of size. In most cases, the measure of size was total district enrollment, but in very small districts a minimum measure of size was used. Each sampled district was asked to update lists of eligible schools according to information on the QED files. Frames of eligible new schools were then constructed separately for both main and long-term trend at each age class, and separate samples of new schools were selected systematically with probability proportional to eligible enrollment using the same sampling rates as for the QED and supplemental nonpublic schools. Thirteen new schools were added to the main samples—one at age class 9, eight at age class 13, and four at age class 17. One new school at age class 13 was added to the long-term trend samples. No new schools were added for age classes 9 and 17. Although new school sampling procedures were applied at these age classes, no new schools were selected since the sampling intervals were large compared to the number of schools eligible for sampling.

In a few PSUs where school refusals were relatively heavy for a particular sample, substitute school selections were made, replacing the refusals (to the extent feasible) with schools from within the same PSU and similar in size, affiliation (public, Catholic, or other nonpublic), grade span, and minority composition. The goal of this procedure was to maintain the student sample sizes needed, while keeping variance and nonresponse bias at acceptable levels. Tables 3-5 and 3-6 show the number of in-scope schools selected, cooperating, and substituted, in the main and long-term trend samples respectively. The participation rates given are based on the original sample of schools (excluding substitutes). School participation rates for long-term trend at age class 13 appear lower compared to those achieved in 1992. The other response rates are comparable for the two years. Note that since the response rates quoted do not include the substitute selections, the potential for nonresponse bias is likely to be a little less than these rates would indicate. This is because the substitute selections were chosen based on their similarity to the initially refusing selections.

Table 3-5

School Sample Sizes, Refusals, and Substitutes for the Main Samples

Status	Age Class 9	Age Class 13	Age Class 17	Total	Public*	Nonpublic**
Selected, in scope	478	712	560	1,750	1,098	652
Refusals	67	103	120	290	194	96
Participation rate of originally selected schools	86%	86%	79%	83%	82%	85%
1992 participation rate	86%	85%	81%	84%	86%	82%
Participating, no eligibles enrolled	15	72	19	106	59	47
Substitutes participating	3	2	2	7	6	1
Final assessed sample	399	539	423	1,361	851	510

*Note: *Public, Bureau of Indian Affairs, and Department of Defense schools*

**Note: *Catholic and other nonpublic schools*

Table 3-6

School Sample Sizes, Refusals, and Substitutes for the Long-Term Trend Samples

Status	Age Class 9	Age Class 13	Age Class 17	Total
Selected, in scope	286	327	244	857
Refusals	38	60	46	144
Participation rate of originally selected schools	87%	82%	81%	83%
1992 participation rate	87%	85%	81%	85%
Participating, no eligibles enrolled	5	28	7	40
Substitutes participating	0	1	4	5
Final assessed sample	243	240	195	678

The considerable numbers of schools selected with no eligible students enrolled resulted primarily from the fact that, for example, for age class 13, some schools with grades 6, 7, or 9, but no grade 8, were sampled. Such schools had a reasonable chance of containing some age 13 students. Often they did have a number of eligible students, but sometimes they had none. Because of the grade structure of schools, this occurred most often for age class 13.

A school characteristics and policies questionnaire was mailed to every sampled school by Westat before the assessment. The Westat supervisor then collected the questionnaires and returned them to ETS. The school characteristics and policies questionnaire is described in Chapter 2. It included information used to refine the estimate of eligible students in the school, which was on the principal's questionnaire in earlier years.

3.4 ASSIGNMENT OF SESSIONS TO SCHOOLS

Sessions were assigned to the selected schools found to be in-scope at the time of session assignment in the following manner. First, the number of sessions per school was established (three for the long-term trend samples, five for the main samples at age class 9, and six for the other main samples). This was the maximum number of sessions that could be administered without creating unduly small session sizes with few eligibles. Thus, in most long-term trend schools, for example, three sessions were conducted. However, schools with fewer than 20 eligibles were asked to conduct only a single session. In the main samples, two different session types were conducted at each age class (reading and U.S. history/geography). For long-term trend, four session types were conducted at age classes 9 and 13 (spiral plus three tape sessions), and three were conducted at age class 17 (spiral plus two tape sessions). Schools could be assigned multiple sessions of the same type (for example, one spiral and two tape sessions in long-term trend, or three reading and three U.S. history/geography sessions in the main samples). Table 3-7 in the following section shows, among other things, the various assessments conducted for each age class and the year to which the long-term trend assessments provide a bridge.

Sessions were assigned to schools with three aims in mind. The first was to distribute students to the different session types across the whole sample for each age class so that the target numbers of assessed students would be achieved. The second was to maximize the number of different session types that were administered within a given selected school, without violating the minimum session sizes discussed above. The third was to give each student an equal chance of being selected for a given session type regardless of the number of sessions conducted in the school. Note that sessions were not explicitly assigned within each PSU as had been done in 1992. The large relative target sample sizes of the session types conducted at each age class made this unnecessary. In fact, each session type was assigned in each PSU, but since sessions were assigned before school cooperation was final, occasionally all session types were not conducted in all PSUs. Adjustments to the sample weights were made for school nonresponse, as discussed in Chapter 10.

3.5 SAMPLING STUDENTS

In the fourth stage of sampling, a consolidated list was prepared for each school of all grade-eligible and age-eligible students for the age class for which the school was selected. A systematic selection of eligible students was made from this list (unless all students were to be assessed) to provide the target sample size. For schools assigned to more than a single session type (the vast majority), students were assigned by Westat district supervisors to one of the various session types using specified procedures.

For each age class, separately for the long-term trend and main samples, maxima were established as to the number of students who would be selected for a given school. In those schools that, according to information on the frame, had fewer eligible students than the established maxima, each eligible student enrolled at the school was selected in the sample for one of the sessions assigned to the school. In other schools, a sample of students was drawn, and then students were assigned to sessions as appropriate. For the main samples, the maximum sample sizes were established in terms of the number of grade- plus age-eligible students—150 at age class 9 and 200 at age classes 13 and 17. For the long-term trend samples, the maximum at each age class was 60 age-eligible students or about 80 grade- plus age-eligible students. Note that the number of students actually selected for assessment in a long-term trend sample school generally fell somewhat below 80, because students who were selected for one of the long-term trend tape-administered sessions and were in the modal grade but not age-eligible were subsequently dropped from the sample.

The sample of students to be selected in each school was derived in the following manner, both for main and for long-term trend samples. On the basis of data obtained from the school characteristics and policies questionnaire (or the sample frame when the questionnaire data were not obtained in time) an estimate of the number of eligible students was established for each school. For the main samples, the estimated number of grade- plus age-eligible students was used; for the long-term trend samples, the number of age-eligible students was used. A Session Assignment Form was generated for each school, showing the line numbers (described below) of the students to be selected, indicating the type of session to be taken by each such student. These line numbers were generated using a sampling interval designed to give the appropriate sample size for each school. Thus the overall sampling interval was 1.0 for schools in which all eligible students were to be assessed. The appropriate sampling interval was specified for schools with larger numbers of eligible students, such as to give the appropriate maximum sample size (described above for each age class) in the case that the school had an enrollment of eligible students exactly equal to that predicted.

If the Westat supervisor found that, when applied to the numbered list of eligible students assembled in the field for each school, the line numbers generated gave rise to a sample in excess of 120 percent of the appropriate maximum sample size limit specified above, he or she called Westat's central office. By use of a personal computer, new line numbers based on the actual number of eligible students were generated and relayed to the supervisor. A similar revision to the line numbers was made in the case of a school with a sampling interval in excess of 1.0, and eligible enrollment less than 80 percent of that initially estimated. In this latter case the sample size was increased to the appropriate level. This procedure gave a suitable compromise between control over the sampling rate within each school and operational autonomy and flexibility for Westat field supervisors. Note that in all cases, sampling intervals were generated in Westat's central office, and stored for use in sample weighting. Supervisors were not required to derive or record within-school sampling rates.

Table 3-7 shows the number of students per school who were assessed for each assessment. Note that, for the various print samples, the number of students assessed per item per school is quite low, even though typically dozens of students were assessed in total in a particular school. Thus the extent of clustering of the sample is in general quite modest, because most sampled schools conducted a few different assessments with a moderate number of students in each, and more importantly because the use of BIB-spiraling in the print-administered sessions greatly alleviated the effects of clustering the samples of students within schools.

3.6 EXCLUDED STUDENTS

Some students selected for the sample were deemed unassessable by school authorities because they had limited English language proficiency, were judged as being mildly mentally retarded (educable), or were functionally disabled. In these cases, school staff completed an IEP/LEP student questionnaire, listing the reason for exclusion and providing some background information (these questionnaires were also completed for IEP/LEP students who were assessed).

Six distinct samples of excluded students were identified. For each age class, there was one sample for the long-term trend assessment and one for the main assessment.

The exclusion criteria for the main samples differed somewhat from those used for the long-term trend samples. The exclusion criteria for the main samples were identical to those used since 1990 and were intended to be somewhat more rigorously defined than those used in the long-term trend samples. (For more details of the exclusion criteria and their implementation, see Chapter 5.) In addition, for age class 17, the excluded students from the long-term trend assessments (with an October-September age definition and modal grade of 11) were drawn from a population different from that of the excluded students in the main assessment (with a calendar-year age definition and modal grade of 12).

For all samples, students were selected for specific sessions, and the school was then asked to identify those to be excluded. Thus only age-eligible students were considered for exclusion from the long-term trend tape-administered sessions, whereas both age- and grade-eligible students were considered for exclusion from the main samples and the print-administered long-term trend samples. The samples of excluded students for the long-term trend samples were weighted in such a way as to account for this procedure appropriately (see Chapter 10).

Table 3-7

Number of Students Per School for Each Assessment

Sample	Sample Code	Assessment	Number of Schools	Mean Number of Students Per Assessment Per School	Mean Number of Students Per Item Per School
Age Class 9 Long-term trend	RW-LTTrend	Print booklets 51-56	217	24.6	4.1 - 8.2*
	MS-LTTrend	Tape booklet 91	130	14.6	14.6
	MS-LTTrend	Tape booklet 92	126	14.9	14.9
	MS-LTTrend	Tape booklet 93	127	14.9	14.9
Age Class 9 Main	Rdg-MainP	Print Reading	321	30.2	7.5
	His-MainP	Print U.S. History	361	20.0	6.7
	Geo-MainP	Print Geography	361	20.0	6.7
Age Class 13 Long-term trend	RW-LTTrend	Print booklets 51-56	212	26.2	4.4 - 8.7*
	MS-LTTrend	Tape booklet 91	125	16.3	16.3
	MS-LTTrend	Tape booklet 92	124	15.8	15.8
	MS-LTTrend	Tape booklet 93	126	16.3	16.3
Age Class 13 Main	Rdg-MainP	Print Reading	464	45.0	5.4-15.1*
	His-MainP	Print U.S. History	452	25.8	5.2
	Geo-MainP	Print Geography	452	20.3	5.1
Age Class 17 Long-term trend	RW-LTTrend	Print booklets 51-56	189	25.6	4.3 - 8.5*
	MS-LTTrend	Tape booklet 84	135	14.0	14.0
	MS-LTTrend	Tape booklet 85	140	13.7	13.7
Age Class 17 Main	Rdg-MainP	Print Reading	388	50.3	5.4 - 15.1*
	His-MainP	Print U.S. History	394	27.5	5.5
	Geo-MainP	Print Geography	394	21.8	7.3

*****Note:** This number varied because some item blocks appeared more than once in the set of booklets used for this sample.*

BEST COPY AVAILABLE

Table 3-8 shows the rates of exclusion for each age class for the long-term trend and main samples. For the main samples, for which nonpublic school students were oversampled by a factor of three and constituted about 16 percent of the student sample, exclusion rates are shown for both public and nonpublic schools. Overall rates for 1992 (when oversampling of nonpublic schools also took place) are shown for comparison. Since 1992, exclusion rates for the main samples appear to have increased, and exclusion rates for long-term trend samples appear to have decreased. Some of the factors that may limit the comparability of the rates across the two years are the different subjects assessed and the fact that substantially more students were classified as IEP in 1994. Somewhat lower rates of exclusion have occurred in the main than in the long-term trend samples, but this is exacerbated by the presence of higher proportions of nonpublic school students in the main samples. The most marked effects, however, are the much higher rates of exclusion in public schools than in nonpublic, and the higher rates of exclusion at lower grades. The former phenomenon is no doubt a function of the greater prevalence of special education and language minority programs in public schools. The higher exclusion rates at lower ages, which occurred also in other years, result from the greater proportion of students at these grades who are excluded for reasons of limited English proficiency. In certain areas of the United States, fourth-grade public-school students whose native language is Spanish are taught predominantly in Spanish, and in these schools a very high proportion of sampled students are excluded.

Table 3-8

Student Exclusion Rates by Age Class and School Type, Unweighted

	1994 Exclusion Rate			1992 Exclusion Rate
Sample	**Public**	**Nonpublic**	**Total***	**Total***
Age Class 9				
Long-term trend	—	—	7.9%	6.0%
Main	7.8%	0.5%	6.6%*	8.8%*
Age Class 13				
Long-term trend	—	—	6.2%	5.4%
Main	5.8%	0.7%	5.1%*	7.4%*
Age Class 17				
Long-term trend	—	—	5.0%	5.6%
Main	5.1%	1.0%	4.5%*	5.9%*

****Note:*** *Somewhat different exclusion criteria were used for the main samples than for the long-term trend samples. Note also that the total rates for the main samples are based on a relatively greater contribution from nonpublic-school students. Nonpublic-school students constitute about 16 percent of the 1994 main samples, 18 percent of the 1992 main samples, 11 percent of the 1994 long-term trend samples, and 11 percent of the 1992 long-term trend samples.*

3.7 STUDENT PARTICIPATION RATES

Table 3-9 summarizes the rates of participation of invited students. The set of invited students consists of the selected students, after removing the excluded students and, in the case of long-term trend samples, removing those students selected for tape-administered sessions who were not age-eligible. For a given session, a makeup session was called for when, for various reasons, more than a tolerable number of invited students failed to attend the originally scheduled session to which they were invited. The participation rates given in the table express the number finally assessed as a percentage of those initially invited in the participating schools. Participation rates are shown for the main and long-term trend samples and for public and nonpublic schools separately in the case of the main samples. Overall participation rates are also shown for comparable samples from the 1992 NAEP assessment. The table shows that student participation rates in 1994 are similar to those experienced in 1992. The rates increased slightly for age class 13 and long-term trend at age class 17, and remained fairly steady for the other samples. At all age classes, the participation rate of nonpublic school students exceeds that of public school students, with the difference, both relative and absolute, increasing with age class. This is in contrast with the levels of school participation, which are higher for public schools.

Table 3-9

Student Participation Rates by Age Class and School Type, Unweighted

	1994 Public		1994 Nonpublic		1994 Combined		
Samples	**Number Invited**	**Participation Rate**	**Number Invited**	**Participation Rate**	**Number Invited**	**Participation Rate***	**1992 Participation Rate***
Age Class 9							
Long-term trend	—	—	—	—	11,672	94.2%	94.0%
Main	21,209	92.9%	4,625	94.8%	25,834	93.2%*	93.4%*
Age Class 13							
Long-term trend	—	—	—	—	12,584	92.2%	90.8%
Main	38,476	90.2%	7,367	95.2%	45,843	91.0%*	88.8%*
Age Class 17							
Long-term trend	—	—	—	—	10,283	84.1%	82.8%
Main	40,755	79.2%	7,243	92.3%	47,998	81.1%*	80.8%*

**Note: The total rates for the 1994 main samples and the 1992 samples are based on a relatively greater contribution of nonpublic-school students than the 1994 long-term trend samples. Nonpublic-school students constitute about 16 percent of the 1994 main samples, 18 percent of the 1992 main samples, 11 percent of the 1994 long-term trend samples, and 11 percent of the 1992 long-term trend samples.*

3.8 OVERALL STUDENT PARTICIPATION RATES

The combined impact of school nonparticipation and student absenteeism from sessions within participating schools is summarized in Table 3-10. The table shows the percentages of students assessed, from among those who would have been assessed if all initially selected schools had participated, and if all invited students had attended either an initial or make-up session. The results show that, consistent with earlier rounds of NAEP, the overall level of participation decreases substantially with the increase in age and grade of the students.

Table 3-10

Overall Participation Rates (School and Student Combined) by Age Class, Unweighted

1994 Samples	Age Class 9	Age Class 13	Age Class 17	Overall
Main Samples				
School participation	86.0%	85.5%	78.6%	83.4%
Student participation	93.2%	91.0%	81.1%	87.5%
Overall student participation	80.2%	77.8%	63.7%	73.0%
Number of participating students	24,090	41,728	38,946	104,764
Long-Term Trend Samples				
School participation	86.7%	81.7%	81.1%	83.2%
Student participation	94.2%	92.2%	84.1%	90.5%
Overall student participation	81.7%	75.3%	68.2%	75.3%
Number of participating students	10,998	11,599	8,653	31,250
Overall				
School participation	86.3%	84.3%	79.4%	83.4%
Student participation	93.6%	91.3%	81.7%	88.2%
Overall student participation	80.8%	77.0%	64.9%	73.6%
Number of participating students	35,088	53,327	47,599	136,014

In past NAEP technical reports, only unweighted participation rates by age class and school type have been presented. However, analysis is typically performed separately by age class and subject, and NCES standards regarding acceptable potentials for bias are expressed in terms of weighted participation rates. Therefore, Tables 3-11 and 3-12 show weighted participation rates by age class and subject for the main and long-term trend samples, respectively. Note that these participation rates are similar for different subjects in the same age class. They are also similar, in general, to the unweighted rates.

The procedures for substituting for nonparticipating schools or imputing for them and the procedures for imputing for absent students were designed (so far as feasible) to reduce the biases resulting from school and student nonparticipation. These procedures are discussed in Chapter 10.

Table 3-11

Weighted Participation Rates by Age Class and Subject, Main Samples

Participation	Reading Print	U.S. History Print	Geography Print
Age Class 9			
School participation	86.1%	86.2%*	86.2%*
Student participation	93.5%	93.7%	93.8%
Overall participation	80.5%	80.8%	80.9%
Age Class 13			
School participation	82.9%	82.3%*	82.3%*
Student participation	91.1%	91.0%	90.9%
Overall participation	75.5%	74.9%	74.8%
Age Class 17			
School participation	76.3%	75.9%*	75.9%*
Student participation	81.9%	81.8%	81.5%
Overall participation	62.5%	62.1%	61.9%

******Note*: *School participation rates are identical for U.S. history and geography within the same age class because both subjects were administered in the same session.*

Table 3-12

Weighted Participation Rates by Age Class and Subject
Long-Term Trend Samples

Participation	Reading/Writing Print	Math/Science Tape
Age Class 9		
School participation	86.7%	87.1%
Student participation	94.1%	94.4%
Overall participation	81.6%	82.2%
Age Class 13		
School participation	79.7%	80.4%
Student participation	91.8%	92.3%
Overall participation	73.2%	74.2%
Age Class 17		
School participation	80.1%	79.5%
Student participation	84.2%	84.8%
Overall participation	67.4%	67.4%

3.9 SAMPLING TEACHERS

The teacher questionnaire was administered to teachers of fourth-grade and eighth-grade students assessed in reading, U.S. history, or geography. Teachers were given the questionnaire if they taught the student the subject in which the student was assessed. The purpose of drawing these samples was not to estimate the attributes of the teacher population, but to estimate the number (proportion) of students whose teachers had various attributes and to correlate student characteristics and performance with the characteristics of their teachers.

The selected teachers were asked to complete a questionnaire concerning themselves and their teaching practices, with specific references to each individual class period containing a student included in the main assessment.

Chapter 4

ASSESSMENT INSTRUMENTS[1]

Stephen Lazer
Educational Testing Service

4.1 INTRODUCTION

In the 1994 assessment, four types of instruments were used to collect data about students, teachers, and schools. Each assessed student received an **assessment booklet** containing both cognitive and background questions. An **IEP/LEP student questionnaire** was completed by school officials for each sampled student who had an Individual Education Plan (IEP) or was classified as Limited English Proficient (LEP), whether or not the students were able to participate in the assessment. The teachers of fourth- and eighth-grade students participating in the assessment were asked to complete a **teacher questionnaire.** A **school characteristics and policies questionnaire** was distributed to each participating school.

In addition, two special studies were conducted. One sample of students at grades 8 and 12 were asked to read specific stories in the *NAEP Reader* and complete the assessment questions. In addition, a small subsample of students who took part in the main U.S. history assessment at grade 8 were asked to take part in a special study of group assessment.

This chapter begins with a discussion of the characteristics of the student booklets used for the 1994 main and long-term trend assessments and how the booklets were assembled. The contents of each booklet and item block is presented in detail in a set of tables. Section 4.4 describes the student, teacher, IEP/LEP, and school questionnaires that were part of the 1994 assessment.

4.2 STUDENT BOOKLETS—MAIN ASSESSMENTS

4.2.1 Reading

Each student assessed in reading received a booklet containing a set of general background questions, reading passages and content questions, a set of subject-specific background questions, and a set of questions about his or her motivation and familiarity with the assessment materials. The passages and content questions were assembled into sections or blocks, each containing either one or two passages and the corresponding questions. Students were given either two 25-minute blocks or one 50-minute block. The overall assessment time for each student was approximately 63 minutes.

---

[1] Stephen Lazer manages assessment development activities for the NAEP program at ETS. He also coordinated the development of the 1994 U.S. history and geography assessments.

BEST COPY AVAILABLE

The assembly of reading blocks into booklets and their subsequent assignment to sampled students was determined by a *partially balanced incomplete block* (PBIB) design with *spiraled* administration. The student booklets contained two five-minute background sections, a one-minute background section, and two 25-minute blocks of reading passages and items according to a PBIB design.[2]

The reading blocks were assigned to booklets in such a way that every 25-minute block within a given purpose for reading (either Reading for Literary Experience, Reading to Gain Information, or Reading to Perform a Task) was paired with every other block measuring the same purpose, but was only paired with one block measuring another purpose for reading. At all grades, every 25-minute block appears in four booklets. At grade 4, each block is paired three times with another block that measures the same purpose for reading. A given block is also paired once with a block measuring the other purpose for reading. At grades 8 and 12, every 25-minute block is paired once with each of the two other blocks that measure the same purpose for reading. A given block is also paired with one block measuring each of the other two purposes for reading. This is the *partially balanced* part of the balanced incomplete block design. Every 50-minute block appears only in one booklet. The PBIB design is equivalent to that used in 1992.

For example, in Table 4-1 block R3 at grade 4 measures Reading for Literary Experience. It is paired with all of the other literary experience blocks (R4, R5, and R9) in booklets 1, 2, and 6. It is also paired with one of the blocks measuring Reading to Gain Information (R8) in booklet 14. Block 3 also occurs twice in the first position (booklets 2 and 6) and twice in the second position (booklets 1 and 14).

The PBIB design for the 1994 national reading assessment was *focused* by subject area, so that students received booklets containing only blocks of reading questions (not U.S. history or geography). The PBIB design also balances the order of presentation of the 25-minute blocks of items—every 25-minute block appears as the first cognitive block in two booklets and as the second cognitive block in two other booklets. This design allows for some reduction of the impact of context and fatigue effects to be measured and reported.

The design used in 1994 required that eight blocks of grade 4 reading items be assembled into 16 booklets. At grades 8 and 12, the 25-minute blocks were assembled into 18 booklets. At grade 8, there were two additional booklets containing 50-minute blocks; at grade 12 there were three of these booklets. Once assembled, the assessment booklets were then *spiraled* and bundled. Spiraling involves interweaving the booklets in a systematic sequence so that each booklet appears an appropriate number of times in the sample. The bundles were designed so that each booklet would appear equally often in each position in a bundle.

The final step in the PBIB-spiraling procedure was the assigning of the booklets to the assessed students. The students within an assessment session were assigned booklets in the order in which the booklets were bundled. Thus, most students in an assessment session received different booklets. In the assessment design, representative and randomly equivalent samples of about 2,500 students responded to each item at a given age/grade level.

---

[2] Booklets containing 50-minute blocks are included in the spiraled administration, but cannot be assembled in the PBIB fashion.

Tables 4-1, 4-3, and 4-5 provide the composition and number of blocks administered in the 1994 reading assessment. Tables 4-2, 4-4, and 4-6 give details of the item blocks used in the main reading assessment, including the number of cognitive and constructed-response items in each block and the booklets in which each block appeared.

Table 4-1
Main Sample Booklet Contents and Number of Booklets Administered
Age 9/Grade 4, Reading

Subject Area	Booklet Number	Common Background Block	Cognitive Blocks*		Subject Area Background Block	Motivation Background Block
Reading	R1	B1	R4(L)	R3(L)	R2	RX
	R2	B1	R3(L)	R5(L)	R2	RX
	R3	B1	R5(L)	R9(L)	R2	RX
	R4	B1	R9(L)	R4(L)	R2	RX
	R5	B1	R4(L)	R5(L)	R2	RX
	R6	B1	R3(L)	R9(L)	R2	RX
	R7	B1	R6(I)	R10(I)	R2	RX
	R8	B1	R10(I)	R7(I)	R2	RX
	R9	B1	R7(I)	R8(I)	R2	RX
	R10	B1	R8(I)	R6(I)	R2	RX
	R11	B1	R6(I)	R7(I)	R2	RX
	R12	B1	R10(I)	R8(I)	R2	RX
	R13	B1	R7(I)	R4(L)	R2	RX
	R14	B1	R8(I)	R3(L)	R2	RX
	R15	B1	R5(L)	R6(I)	R2	RX
	R16	B1	R9(L)	R10(I)	R2	RX

**Note: The letter "L" denotes a block designed to measure Reading for Literary Experience and "I" indicates Reading to Gain Information.*

Table 4-2

Cognitive and Noncognitive Block Information

Age 9/Grade 4, Reading

Block	Type*	Total Number of Items	Number of Multiple-Choice Items	Number of Constructed-Response Items	Booklets Containing Block
B1	Common Background	22	22	0	1 - 16
R2	Reading Background	15	15	0	1 - 16
RX	Reading Motivation	5	5	0	1 - 16
R3	Reading Cognitive (L)	11	6	5	1, 2, 6, 14
R4	Reading Cognitive (L)	12	5	7	1, 4, 5, 13
R5	Reading Cognitive (L)	11	7	4	2, 3, 5, 15
R6	Reading Cognitive (I)	10	5	5	7, 10, 11, 15
R7	Reading Cognitive (I)	10	4	6	8, 9, 10, 11
R8	Reading Cognitive (I)	9	3	6	9, 10, 12, 14
R9	Reading Cognitive (L)	9	3	6	3, 4, 6, 16
R10	Reading Cognitive (I)	12	6	6	7, 8, 12, 16
	Total Cognitive	**84**	**39**	**45**	

**Note*: At grade 4, all reading cognitive blocks are 25-minutes long. The letter "L" denotes a block designed to measure Reading for Literary Experience and "I" indicates Reading to Gain Information.*

Table 4-3

Main Sample Booklet Contents and Number of Booklets Administered
Age 13/Grade 8, Reading

Subject Area	Booklet Number	Common Background Block	Cognitive Blocks*		Subject Area Background Block	Motivation Background Block
Reading	R1	B1	R3	R4	RB	RX
	R2	B1	R4	R5	RB	RX
	R3	B1	R5	R3	RB	RX
	R4	B1	R6	R8	RB	RX
	R5	B1	R8	R7	RB	RX
	R6	B1	R7	R6	RB	RX
	R7	B1	R10	R9	RB	RX
	R8	B1	R9	R11	RB	RX
	R9	B1	R11	R10	RB	RX
	R10	B1	R3	R8	RB	RX
	R11	B1	R7	R4	RB	RX
	R12	B1	R5	R6	RB	RX
	R13	B1	R6	R9	RB	RX
	R14	B1	R8	R11	RB	RX
	R15	B1	R10	R7	RB	RX
	R16	B1	R4	R10	RB	RX
	R17	B1	R9	R5	RB	RX
	R18	B1	R11	R3	RB	RX
	R19N	B1	R12		RB	RX
	R20N	B1	R12A		RB	RX
	R21	B1	R13		RB	RX

*Note: Blocks R12 and R12A require the NAEP Reader; blocks R12 and R13 are 50-minute blocks. Block 12A is a forced-choice version of the NAEP Reader block.

Table 4-4

Cognitive and Noncognitive Block Information
Age 13/Grade 8, Reading

Block	Type*	Total Number of Items	Number of Multiple-Choice Items	Number of Constructed-Response Items	Booklets Containing Block
B1	Common Background	21	21	0	1 - 21
R2	Reading Background	25	25	0	1 - 21
RX	Reading Motivation	5	5	0	1 - 21
R3	Reading Cognitive (L)	11	4	7	1, 3, 10, 18
R4	Reading Cognitive (L)	8	1	7	1, 2, 11, 16
R5	Reading Cognitive (L)	11	7	4	2, 3, 12, 17
R6	Reading Cognitive (I)	12	5	7	4, 6, 12, 13
R7	Reading Cognitive (I)	13	6	7	5, 6, 11, 15
R8	Reading Cognitive (I)	9	3	6	4, 5, 10, 14
R9	Reading Cognitive (T)	9	4	5	7, 8, 13, 17
R10	Reading Cognitive (T)	12	4	8	7, 9, 15, 16
R11	Reading Cognitive (T)	12	3	9	8, 9, 14, 18
R12	Reading Cognitive (L)	12	0	12	19, 20
R13	Reading Cognitive (I)	13	4	9	21
	Total Cognitive	**122**	**41**	**81**	

**Note: All reading cognitive blocks are 25-minutes long except R12 and R13, which are 50 minutes long. The letter "L" denotes a block designed to measure Reading for Literary Experience, "I" indicates Reading to Gain Information, and "T" denotes Reading to Perform a Task.*

Table 4-5

Main Sample Booklet Contents and Number of Booklets Administered
Age 17/Grade 12, Reading

Subject Area	Booklet Number	Common Background Block	Cognitive Blocks*		Subject Area Background Block	Motivation Background Block
Reading	R1	B1	R3	R4	RB	RX
	R2	B1	R4	R5	RB	RX
	R3	B1	R5	R3	RB	RX
	R4	B1	R6	R7	RB	RX
	R5	B1	R7	R8	RB	RX
	R6	B1	R8	R6	RB	RX
	R7	B1	R10	R9	RB	RX
	R8	B1	R9	R11	RB	RX
	R9	B1	R11	R10	RB	RX
	R10	B1	R3	R7	RB	RX
	R11	B1	R8	R4	RB	RX
	R12	B1	R5	R6	RB	RX
	R13	B1	R6	R9	RB	RX
	R14	B1	R7	R11	RB	RX
	R15	B1	R10	R8	RB	RX
	R16	B1	R4	R10	RB	RX
	R17	B1	R9	R5	RB	RX
	R18	B1	R11	R3	RB	RX
	R19N	B1	R12		RB	RX
	R20N	B1	R12A		RB	RX
	R21	B1	R13		RB	RX
	R22	B1	R14		RB	RX

**Note*:* *Blocks R12 and R12A require the NAEP Reader; blocks R12, R13, and R14 are 50-minute blocks. Block 12A is a forced-choice version of the NAEP Reader block.*

Table 4-6

Cognitive and Noncognitive Block Information
Age 17/Grade 12, Reading

Block	Type*	Total Number of Items	Number of Multiple-Choice Items	Number of Constructed-Response Items	Booklets Containing Block
B1	Common Background	21	21	0	1 - 22
R2	Reading Background	25	25	0	1 - 22
RX	Reading Motivation	5	5	0	1 - 22
R3	Reading Cognitive (L)	11	4	7	1, 3, 10, 18
R4	Reading Cognitive (L)	9	3	6	1, 2, 11, 16
R5	Reading Cognitive (L)	8	1	7	2, 3, 12, 17
R6	Reading Cognitive (I)	12	5	7	4, 6, 12, 13
R7	Reading Cognitive (I)	12	5	7	4, 5, 10, 14
R8	Reading Cognitive (I)	8	1	7	5, 6, 11, 15
R9	Reading Cognitive (T)	9	4	5	7, 8, 13, 17
R10	Reading Cognitive (T)	12	4	8	7, 9, 15, 16
R11	Reading Cognitive (T)	15	7	8	8, 9, 14, 18
R12	Reading Cognitive (L)	12	0	12	19, 20
R13	Reading Cognitive (I)	16	10	6	21
R14	Reading Cognitive (I)	8	0	8	22
	Total Cognitive	**132**	**44**	**88**	

__Note__: All reading cognitive blocks are 25-minutes long except R12, R13, and R14, which are 50 minutes long. The letter "L" denotes a block designed to measure Reading for Literary Experience, "I" indicates Reading to Gain Information, and "T" denotes Reading to Perform a Task.

4.2.2 U.S. History

Each student assessed in U.S. history received a booklet containing a set of general background questions, content questions, a set of subject-specific background questions, and a set of questions about his or her motivation and familiarity with the assessment materials. The content questions were assembled into sections or blocks, each containing a range of questions covering the four U.S. history themes. There were two types of content blocks on the assessment: 25-minute blocks that surveyed a range of student knowledge in U.S. history and, at grades 8 and 12, a 50-minute thematic block that focused on a specific history topic. Students were given either two 25-minute blocks or one 50-minute block. The overall assessment time for each student was approximately 63 minutes.

Each student assessment booklet contained two five-minute background sections, one three-minute motivation section, and either one 50-minute or two 25-minute U.S. history blocks. The assembly of U.S. history blocks into booklets and their subsequent assignment to sampled students was determined by a balanced incomplete block (BIB) design with spiraled administration. The U.S. history blocks were assigned to booklets in such a way that every 25-minute block was paired with every other block. At grade 4, every block appears in six booklets. At grades 8 and 12, every 25-minute block appears in eight booklets. Every 50-minute block appears only in one booklet.

The BIB design for the 1994 U.S. history assessment was focused by subject area, so that students received booklets containing only blocks of U.S. history questions (not reading or geography). The focused-BIB design also balances the order of presentation of the 25-minute blocks of items so that every 25-minute block appears as the first cognitive block and as the second cognitive block an equal number of times (three times in each position at grade 4 and four times in each position at the older grades). This design allows for some reduction of the impact of context and fatigue effects to be measured and reported.

The design used in 1994 required that six blocks of grade 4 U.S. history items be assembled into 18 booklets. At grades 8 and 12, the eight 25-minute blocks were assembled into 32 booklets and there was one additional booklet containing a 50-minute block. Once assembled, the assessment booklets were then spiraled and bundled. Spiraling involves interweaving the booklets in a systematic sequence so that each booklet appears an appropriate number of times in the sample. The bundles were designed so that each booklet would appear equally often in each position in a bundle. The exception to this rule was in the case of booklets containing 50-minute blocks that appeared more often to ensure that an adequate number of students took each item.

The final step in the BIB-spiraling procedure was the assigning of the booklets to the assessed students. The students within an assessment session were assigned booklets in the order in which the booklets were bundled. Thus, most students in an assessment session received different booklets. In the assessment design, representative and randomly equivalent samples of about 2,500 students responded to each item at a given age/grade level.

Tables 4-7 and 4-9 provide the composition and number of blocks administered in the 1994 U.S. history assessment. Tables 4-8, 4-10, and 4-11 give details of the item blocks used in the main U.S. history assessment, including the number of cognitive and constructed-response items in each block and the booklets in which each block appeared.

Table 4-7
Main Sample Booklet Contents and Number of Booklets Administered
Age 9/Grade 4, U.S. History

Subject Area	Booklet Number	Common Background Block	Cognitive Blocks		Subject Area Background Block	Motivation Background Block
U.S. History	H101	B1	H3	H4	H2	HX
	H102	B1	H4	H5	H2	HX
	H103	B1	H5	H6	H2	HX
	H104	B1	H6	H7	H2	HX
	H105	B1	H7	H8	H2	HX
	H106	B1	H8	H3	H2	HX
	H107	B1	H3	H5	H2	HX
	H108	B1	H4	H6	H2	HX
	H109	B1	H5	H7	H2	HX
	H110	B1	H6	H8	H2	HX
	H111	B1	H7	H3	H2	HX
	H112	B1	H8	H4	H2	HX
	H113	B1	H3	H6	H2	HX
	H114	B1	H4	H7	H2	HX
	H115	B1	H5	H8	H2	HX
	H116	B1	H6	H3	H2	HX
	H117	B1	H7	H4	H2	HX
	H118	B1	H8	H5	H2	HX

Table 4-8

Cognitive and Noncognitive Block Information
Age 9/Grade 4, U.S. History

Block	Type*	Total Number of Items	Number of Multiple-Choice Items	Number of Constructed-Response Items	Booklets Containing Block
B1	Common Background	22	22	0	101 - 118
H2	U.S. History Background	17	17	0	101 - 118
HX	U.S. History Motivation	5	5	0	101 - 118
H3	U.S. History Cognitive	15	10	5	101, 106, 107, 111, 113, 116
H4	U.S. History Cognitive	15	9	6	101, 102, 108, 112, 114, 117
H5	U.S. History Cognitive	16	11	5	102, 103, 107, 109, 115, 118
H6	U.S. History Cognitive	15	10	5	103, 104, 108, 110, 113, 116
H7	U.S. History Cognitive	16	11	5	104, 105, 109, 111, 114, 117
H8	U.S. History Cognitive	17	11	6	105, 106, 110, 112, 115, 118
	Total Cognitive	**94**	**62**	**32**	

**Note:* *At grade 4, all cognitive blocks are 25-minutes long.*

Table 4-9

Main Sample Booklet Contents and Number of Booklets Administered
Age 13/Grade 8 and Age 17/Grade 12, U.S. History

Subject Area	Booklet Number	Common Background Block	Cognitive Blocks		Subject Area Background Block	Motivation Background Block
U.S. History	H101	B1	H3	H4	H2	HX
	H102	B1	H4	H5	H2	HX
	H103	B1	H5	H6	H2	HX
	H104	B1	H6	H7	H2	HX
	H105	B1	H7	H8	H2	HX
	H106	B1	H8	H9	H2	HX
	H107	B1	H9	H10	H2	HX
	H108	B1	H10	H3	H2	HX
	H109	B1	H3	H5	H2	HX
	H110	B1	H4	H6	H2	HX
	H111	B1	H5	H7	H2	HX
	H112	B1	H6	H8	H2	HX
	H113	B1	H7	H9	H2	HX
	H114	B1	H8	H10	H2	HX
	H115	B1	H9	H3	H2	HX
	H116	B1	H10	H4	H2	HX
	H117	B1	H3	H6	H2	HX
	H118	B1	H4	H7	H2	HX
	H119	B1	H5	H8	H2	HX
	H120	B1	H6	H9	H2	HX
	H121	B1	H7	H10	H2	HX
	H122	B1	H8	H3	H2	HX
	H123	B1	H9	H4	H2	HX
	H124	B1	H10	H5	H2	HX
	H125	B1	H3	H7	H2	HX
	H126	B1	H4	H8	H2	HX
	H127	B1	H5	H9	H2	HX
	H128	B1	H6	H10	H2	HX
	H129	B1	H7	H3	H2	HX
	H130	B1	H8	H4	H2	HX
	H131	B1	H9	H5	H2	HX
	H132	B1	H10	H6	H2	HX
	H133	B1	H11*		H2	HX

**Note: Block H11 is a 50-minute theme block.*

Table 4-10

Cognitive and Noncognitive Block Information
Age 13/Grade 8, U.S. History

Block	Type*	Total Number of Items	Number of Multiple-Choice Items	Number of Constructed-Response Items	Booklets Containing Block
B1	Common Background	21	21	0	101 - 133
H2	U.S. History Background	23	23	0	101 - 133
HX	U.S. History Motivation	5	5	0	101 - 133
H3	U.S. History Cognitive	17	12	5	101, 108, 109, 115, 117, 122, 125, 129
H4	U.S. History Cognitive	17	12	5	101, 102, 110, 116, 118, 123, 126, 130
H5	U.S. History Cognitive	17	12	5	102, 103, 109, 111, 119, 124, 127, 131
H6	U.S. History Cognitive	17	12	5	103, 104, 110, 112, 117, 120, 128, 132
H7	U.S. History Cognitive	16	11	5	104, 105, 111, 113, 118, 121, 125, 129
H8	U.S. History Cognitive	17	11	6	105, 106, 112, 114, 119, 122, 126, 130
H9	U.S. History Cognitive	17	13	4	106, 107, 113, 115, 120, 123, 127, 131
H10	U.S. History Cognitive	17	12	5	107, 108, 114, 116, 121, 124, 128, 132
H11	U.S. History Cognitive	13	6	7	133
	Total Cognitive	**148**	**101**	**47**	

*****Note:*** *At grade 8, all U.S. history cognitive blocks are 25-minutes long, except H11, which is a 50-minute thematic block.*

Table 4-11

Cognitive and Noncognitive Block Information
Age 17/Grade 12, U.S. History

Block	Type*	Total Number of Items	Number of Multiple-Choice Items	Number of Constructed-Response Items	Booklets Containing Block
B1	Common Background	30	30	0	101 - 133
H2	U.S. History Background	25	25	0	101 - 133
HX	U.S. History Motivation	5	5	0	101 - 133
H3	U.S. History Cognitive	17	12	5	101, 108, 109, 115, 117, 122, 125, 129
H4	U.S. History Cognitive	18	13	5	101, 102, 110, 116, 118, 123, 126, 130
H5	U.S. History Cognitive	17	10	7	102, 103, 109, 111, 119, 124, 127, 131
H6	U.S. History Cognitive	17	12	5	103, 104, 110, 112, 117, 120, 128, 132
H7	U.S. History Cognitive	17	12	5	104, 105, 111, 113, 118, 121, 125, 129
H8	U.S. History Cognitive	17	12	5	105, 106, 112, 114, 119, 122, 126, 130
H9	U.S. History Cognitive	17	13	4	106, 107, 113, 115, 120, 123, 127, 131
H10	U.S. History Cognitive	17	12	5	107, 108, 114, 116, 121, 124, 128, 132
H11	U.S. History Cognitive	19	8	11	133
	Total Cognitive	**156**	**104**	**52**	

**Note*: *At grade 12, all U.S. history cognitive blocks are 25-minutes long, except H11, which is a 50-minute thematic block.*

4.2.3 Geography

Each student assessed in geography received a booklet containing a five-minute section of general background questions, a five-minute section of subject-specific background questions, a one-minute section of questions about his or her motivation and familiarity with the assessment materials, and content questions. The content questions were assembled into sections or blocks, each containing a range of questions covering the three geography content areas.[3] There were two types of content blocks on the assessment: 25-minute blocks that surveyed a range of student knowledge in geography and, at grades 8 and 12, a 50-minute thematic block that focused on a specific geography topic. Students were given either two 25-minute blocks or one 50-minute block. The overall assessment time for each student was approximately 63 minutes.

The assembly of geography blocks into booklets and their subsequent assignment to sampled students was determined by a PBIB design with spiraled administration. The geography blocks were assigned to booklets in such a way that every 25-minute block was paired with every other block. At each grade, every 25-minute block appears in six booklets. At grades 8 and 12, the 50-minute blocks appear in only one booklet.

The BIB design for the 1994 geography assessment was focused by subject area, so that students received booklets containing only blocks of geography questions (not reading or U.S. history). The focused-BIB design also balances the order of presentation of the 25-minute blocks of items so that every 25-minute block appears as the first cognitive block and as the second cognitive block an equal number of times. This design allows for some reduction of the impact of context and fatigue effects to be measured and reported.

The design used in 1994 required that the six 25-minute blocks of geography items at each grade be assembled into 18 booklets. At grades 8 and 12, there was one additional booklets containing a 50-minute block. Once assembled, the assessment booklets were then spiraled and bundled. Spiraling involves interweaving the booklets in a systematic sequence so that each booklet appears an appropriate number of times in the sample. The bundles were designed so that each booklet would appear equally often in each position in a bundle.

The final step in the BIB-spiraling procedure was the assigning of the booklets to the assessed students. The students within an assessment session were assigned booklets in the order in which the booklets were bundled. Thus, most students in an assessment session received different booklets. In the assessment design, representative and randomly equivalent samples of about 2,500 students responded to each item at a given age/grade level.

Tables 4-12 and 4-14 provide the composition and number of blocks administered in the 1994 geography assessment. Tables 4-13, 4-15, and 4-16 give details of the item blocks used in the main geography assessment, including the number of cognitive and constructed-response items in each block and the booklets in which each block appeared.

---

[3] While every attempt was made to ensure that each block covered all content areas, some of the blocks, especially the thematic and atlas-based blocks, did not lend themselves to broad content coverage.

Table 4-12

Main Sample Booklet Contents and Number of Booklets Administered
Age 9/Grade 4, Geography

Subject Area	Booklet Number	Common Background Block	Cognitive Blocks*		Subject Area Background Block	Motivation Background Block
Geography	G31	B1	G3	G4	G2	GX
	G32	B1	G4	G5	G2	GX
	G33	B1	G5	G6	G2	GX
	G34	B1	G6	G7	G2	GX
	G35	B1	G7	G8	G2	GX
	G36	B1	G8	G3	G2	GX
	G37	B1	G3	G5	G2	GX
	G38	B1	G4	G6	G2	GX
	G39	B1	G5	G7	G2	GX
	G40	B1	G6	G8	G2	GX
	G41	B1	G7	G3	G2	GX
	G42	B1	G8	G4	G2	GX
	G43	B1	G3	G6	G2	GX
	G44	B1	G4	G7	G2	GX
	G45	B1	G5	G8	G2	GX
	G46	B1	G6	G3	G2	GX
	G47	B1	G7	G4	G2	GX
	G48	B1	G8	G5	G2	GX

****Note**:* *On block G3 students use colored pencils, and on block G7 students use an atlas. All blocks are 25-minutes long.*

Table 4-13

Cognitive and Noncognitive Block Information
Age 9/Grade 4, Geography

Block	Type*	Total Number of Items	Number of Multiple-Choice Items	Number of Constructed-Response Items	Booklets Containing Block
B1	Common Background	22	22	0	31 - 48
G2	Geography Background	22	22	0	31 - 48
GX	Geography Motivation	5	5	0	31 - 48
G3	Geography Cognitive	14	8	6	31, 36, 37, 41, 43, 46
G4	Geography Cognitive	15	10	5	31, 32, 38, 42, 44, 47
G5	Geography Cognitive	15	11	4	32, 33, 37, 39, 45, 48
G6	Geography Cognitive	15	9	6	33, 34, 38, 40, 43, 46
G7	Geography Cognitive	14	9	5	34, 35, 39, 41, 44, 47
G8	Geography Cognitive	17	12	5	35, 36, 40, 42, 45, 48
	Total Cognitive	**90**	**59**	**31**	

Note: *At grade 4, all cognitive blocks are 25-minutes long. On block G3 students use colored pencils, while on block G7 students use an atlas.*

Table 4-14

Main Sample Booklet Contents and Number of Booklets Administered
Age 13/Grade 8 and Age 17/Grade 12, Geography

Subject Area	Booklet Number	Common Background Block	Cognitive Blocks*		Subject Area Background Block	Motivation Background Block
Geography	G31	B1	G3	G4	G2	GX
	G32	B1	G4	G5	G2	GX
	G33	B1	G5	G6	G2	GX
	G34	B1	G6	G7	G2	GX
	G35	B1	G7	G8	G2	GX
	G36	B1	G8	G3	G2	GX
	G37	B1	G3	G5	G2	GX
	G38	B1	G4	G6	G2	GX
	G39	B1	G5	G7	G2	GX
	G40	B1	G6	G8	G2	GX
	G41	B1	G7	G3	G2	GX
	G42	B1	G8	G4	G2	GX
	G43	B1	G3	G6	G2	GX
	G44	B1	G4	G7	G2	GX
	G45	B1	G5	G8	G2	GX
	G46	B1	G6	G3	G2	GX
	G47	B1	G7	G4	G2	GX
	G48	B1	G8	G5	G2	GX
	G49	B1	G9		G2	GX

__Note__: On block G7 students use an atlas. All blocks are 25-minutes long except G9, which lasts 50 minutes.

Table 4-15

Cognitive and Noncognitive Block Information
Age 13/Grade 8, Geography

Block	Type*	Total Number of Items	Number of Multiple-Choice Items	Number of Constructed-Response Items	Booklets Containing Block
B1	Common Background	21	21	0	31 - 49
G2	Geography Background	25	25	0	31 - 49
GX	Geography Motivation	5	5	0	31 - 49
G3	Geography Cognitive	17	12	5	31, 36, 37, 41, 43, 46
G4	Geography Cognitive	17	12	5	31, 32, 38, 42, 44, 47
G5	Geography Cognitive	18	12	6	32, 33, 37, 39, 45, 48
G6	Geography Cognitive	16	11	5	33, 34, 38, 40, 43, 46
G7	Geography Cognitive	14	9	5	34, 35, 39, 41, 44, 47
G8	Geography Cognitive	17	12	5	35, 36, 40, 42, 45, 48
G9	Geography Cognitive	26	16	10	49
	Total Cognitive	**125**	**84**	**41**	

**Note:* *On block G7 students use an atlas. All blocks are 25-minutes long except G9, which lasts 50 minutes.*

BEST COPY AVAILABLE

Table 4-16

Cognitive and Noncognitive Block Information
Age 17/Grade 12, Geography

Block	Type*	Total Number of Items	Number of Multiple-Choice Items	Number of Constructed-Response Items	Booklets Containing Block
B1	Common Background	30	30	0	31 - 49
G2	Geography Background	29	29	0	31 - 49
GX	Geography Motivation	5	5	0	31 - 49
G3	Geography Cognitive	17	12	5	31, 36, 37, 41, 43, 46
G4	Geography Cognitive	17	12	5	31, 32, 38, 42, 44, 47
G5	Geography Cognitive	18	12	6	32, 33, 37, 39, 45, 48
G6	Geography Cognitive	16	11	5	33, 34, 38, 40, 43, 46
G7	Geography Cognitive	15	11	4	34, 35, 39, 41, 44, 47
G8	Geography Cognitive	17	11	6	35, 36, 40, 42, 45, 48
G9	Geography Cognitive	23	16	7	49
	Total Cognitive	**123**	**85**	**38**	

*Note: *On block G7 students use an atlas. All blocks are 25-minutes long except G9, which lasts 50 minutes.*

4.3 STUDENT BOOKLETS—LONG-TERM TREND ASSESSMENTS

There were several long-term trend samples in the 1994 assessment (see Chapter 1), each of which required the use of special booklets. Tables 4-17, 4-18, and 4-19 summarize the contents of each trend assessment booklet and show how many of each booklet were administered. Tables 4-20, 4-21, and 4-22 give details of the item blocks used in the long-term trend assessments, including the number of cognitive and constructed-response items in each block and the booklets in which each block appeared.

Reading and Writing Long-Term Trend. Six booklets (numbered 51 to 56) containing reading and writing items were administered to each age class. These booklets were identical to booklets used in previous assessments of reading and writing and were spiraled for administration. Each booklet consisted of a common background block (BZ) and three cognitive blocks (at least one reading block and at least one writing block). In addition to cognitive items, the cognitive blocks also contained subject-related background questions.

Mathematics and Science Long-Term Trend. Three booklets (91, 92, and 93) at ages 9 and 13 and two booklets (84 and 85) at age 17, containing mathematics and science items, were identical to those used in previous assessments to measure trends. Each booklet contained a common background block (C1 or BZ) and three cognitive blocks. At ages 9 and 13, these booklets contained one reading block (R1, R2, or R3), one mathematics block (M1, M2, or M3) and one science block (S1, S2, or S3). At age 17, each booklet contained at least one mathematics block (M1 to M3) and at least one science block (S1 - S3). Mathematics block M3 contained items that required the use of a calculator. All cognitive blocks also contained subject-related background questions.

Table 4-17

Long-Term Trend Sample Booklet Contents and Number of Booklets Administered
Age Class 9

Subject Area	Booklet Number	Common Background Block	Subject Area Background Block	Cognitive Blocks			Number of Booklets Administered
Reading and Writing	51	BZ	*	BC	BL	BQ	1,186
	52	BZ	*	BH	BE	BR	1,165
	53	BZ	*	BC	BK	BJ	1,178
	54	BZ	*	BG	BO	BE	1,180
	55	BZ	*	BM	BG	BN	1,169
	56	BZ	*	BV	BR		1,184
Mathematics and Science	91	C1	*	R1	M1	S1	2,388
	92	C1	*	S2	R2	M3**	2,512
	93	C1	*	M2	S3	R3	2,435

***Note**: *Subject area background questions are included in cognitive blocks for this booklet.*
****Note**: *Calculator needed for this block.*

Table 4-18

Long-Term Trend Sample Booklet Contents and Number of Booklets Administered
Age Class 13

Subject Area	Booklet Number	Common Background Block	Subject Area Background Block	Cognitive Blocks			Number of Booklets Administered
Reading and Writing	51	BZ	*	BM	BK	BD	919
	52	BZ	*	BC	BL	BQ	906
	53	BZ	*	BH	BE	BR	923
	54	BZ	*	BN	BC	BD	905
	55	BZ	*	BG	BO	BE	928
	56	BZ	*	BG	BJ	BP	933
Mathematics and Science	91	C1	*	R1	M1	S1	1,928
	92	C1	*	S2	R2	M3**	1,976
	93	C1	*	M2	S3	R3	2,005

***Note**: *Subject area background questions are included in cognitive blocks for this booklet.*
****Note**: *Calculator needed for this block.*

Table 4-19

Long-Term Trend Sample Booklet Contents and Number of Booklets Administered
Age Class 17

Subject Area	Booklet Number	Common Background Block	Subject Area Background Block	Cognitive Blocks			Number of Booklets Administered
Reading and Writing	51	BZ	*	BM	BK	BD	927
	52	BZ	*	BC	BL	BQ	924
	53	BZ	*	BH	BE	BR	917
	54	BZ	*	BN	BC	BD	951
	55	BZ	*	BG	BO	BE	939
	56	BZ	*	BG	BJ	BP	911
Mathematics and Science	84	C1	*	M1	M2	S3	2,207
	85	C1	*	S1	S2	M3**	2,152

***Note**: *Subject area background questions are included in cognitive blocks for this booklet.*
****Note**: *Calculator needed for this block.*

Table 4-20

Long-Term Trend Sample Block Information, Age Class 9

Block	Type	Total Number of Items	Number of Cognitive Items	Number of Open-Ended Items		Booklets Containing Block
				Cognitive	Noncognitive	
BZ	Common Background	37	0	0	1	51 - 56
C1	Common Background	28	0	0	0	91 - 93
BC	Writing Background/Cognitive	23	1	1	0	51, 53
BE	Writing Background/Cognitive	11	2	2	0	52, 54
BG	Writing Background/Cognitive	8	2	2	0	54, 55
BH	Reading Background/Cognitive	15	11	1	0	52
BJ	Reading Background/Cognitive	24	13	1	0	53
BK	Reading Background/Cognitive	19	11	0	0	53
BL	Reading Background/Cognitive	26	7	1	1	51
BM	Reading Background/Cognitive	16	12	1	0	55
BN	Reading Background/Cognitive	25	14	1	0	55
BO	Reading Background/Cognitive	22	11	0	0	54
BQ	Reading Background/Cognitive	21	12	0	0	51
BR	Reading Background/Cognitive	16	12	0	0	52, 56
BV	Reading and Writing Background/Cognitive	36	7 Rd. 1 Wr.	1 Rd. 1 Wr.	0	56
R1	Reading Background/Cognitive	20	9	0	0	91
R2	Reading Background/Cognitive	20	11	0	0	92
R3	Reading Background/Cognitive	17	10	1	0	93
M1	Mathematics Background/Cognitive	26	26	9	0	91
M2	Mathematics Background/Cognitive	26	26	9	0	93
M3	Mathematics Background/Cognitive (Calc.)	19	16	10	0	92
S1	Science Background/Cognitive	23	18	0	0	91
S2	Science Background/Cognitive	25	25	0	0	92
S3	Science Background/Cognitive	31	20	0	0	93

Table 4-21

Long-Term Trend Sample Block Information, Age Class 13

Block	Type	Total Number of Items	Number of Cognitive Items	Number of Open-Ended Items		Booklets Containing Block
				Cognitive	Noncognitive	
BZ	Common Background	37	0	0	1	51 - 56
C1	Common Background	30	0	0	0	91 - 93
BC	Writing Background/Cognitive	23	1	1	0	52, 54
BD	Writing Background/Cognitive	25	1	1	0	51, 54
BE	Writing Background/Cognitive	11	2	2	0	53, 55
BG	Writing Background/Cognitive	8	2	2	0	55, 56
BH	Reading Background/Cognitive	18	13	1	1	53
BJ	Reading Background/Cognitive	24	14	2	0	56
BK	Reading Background/Cognitive	17	9	1	0	51
BL	Reading Background/Cognitive	27	6	1	1	52
BM	Reading Background/Cognitive	16	12	1	0	51
BN	Reading Background/Cognitive	23	12	1	0	54
BO	Reading Background/Cognitive	21	10	2	0	55
BP	Reading Background/Cognitive	15	9	1	0	55
BQ	Reading Background/Cognitive	23	17	0	0	52
BR	Reading Background/Cognitive	19	15	0	0	53
R1	Reading Background/Cognitive	31	12	1	0	91
R2	Reading Background/Cognitive	19	10	0	0	92
R3	Reading Background/Cognitive	28	13	0	0	93
M1	Mathematics Background/Cognitive	51	37	9	0	91
M2	Mathematics Background/Cognitive	44	37	8	0	93
M3	Mathematics Background/Cognitive (Calc.)	32	24	10	0	92
S1	Science Background/Cognitive	36	25	0	0	91
S2	Science Background/Cognitive	40	27	0	0	92
S3	Science Background/Cognitive	36	27	0	0	93

Table 4-22

Long-Term Trend Sample Block Information, Age Class 17

Block	Type	Total Number of Items	Number of Cognitive Items	Number of Open-Ended Items		Booklets Containing Block
				Cognitive	Noncognitive	
BZ	Common Background	48	0	0	1	51 - 56
C1	Common Background	48	0	0	0	84, 85
BC	Writing Background/Cognitive	23	1	1	0	52, 54
BD	Writing Background/Cognitive	25	1	1	0	51, 54
BE	Writing Background/Cognitive	11	2	2	0	53, 55
BG	Writing Background/Cognitive	8	2	2	0	55, 56
BH	Reading Background/Cognitive	19	13	1	2	53
BJ	Reading Background/Cognitive	17	6	2	1	56
BK	Reading Background/Cognitive	17	9	1	0	51
BL	Reading Background/Cognitive	32	6	1	2	52
BM	Reading Background/Cognitive	16	12	1	0	51
BN	Reading Background/Cognitive	32	12	1	1	54
BO	Reading Background/Cognitive	24	13	1	0	55
BP	Reading Background/Cognitive	25	11	1	0	56
BQ	Reading Background/Cognitive	17	11	1	0	52
BR	Reading Background/Cognitive	20	9	0	0	53
M1	Mathematics Background/Cognitive	49	35	10	0	84
M2	Mathematics Background/Cognitive	49	35	5	0	84
M3	Mathematics Background/Cognitive (Calc.)	35	24	14	0	85
S1	Science Background/Cognitive	38	27	0	0	85
S2	Science Background/Cognitive	41	32	0	0	85
S3	Science Background/Cognitive	32	23	0	0	84

4.4 STUDENT, TEACHER, AND SCHOOL QUESTIONNAIRES

4.4.1 Student Questionnaires

Each booklet in the main assessment included three student background questionnaires. The first, consisting of general background questions, included questions about race/ethnicity, mother's and father's level of education, reading materials in the home, homework, attendance, academic expectations, and which parents lived at home. The second, consisting of subject-area background questions, included questions about instructional activities, courses taken, use of specialized resources such as calculators in mathematics class, and views on the utility and value of the subject matter. Students were given five minutes to complete each of these questionnaires, with the exception of the fourth graders, who were given more time because the items in the general questionnaire were read aloud for them. The third questionnaire followed the three cognitive blocks and contained five questions about students' motivation to do well on the assessment, their perceptions concerning the difficulty of the assessment, and their familiarity with types of questions included.

The student questionnaires are described in detail in Chapter 2.

4.4.2 IEP/LEP Student Questionnaires

The IEP/LEP Student Questionnaire was completed by the teachers of those students who were selected to participate in the assessment sample who had Individualized Educational Plans (IEP) or were classified as Limited English Proficient (LEP). The questionnaire was completed for all IEP or LEP students, whether or not they actually participated in the assessment. This questionnaire asked about the nature of the student's disability and the special programs in which the student participated.

Schools were permitted to exclude certain students from the assessment. The same exclusion criteria and rules used in the national assessment were also applied to the Trial State Assessment. Although the intent was to assess all sampled students, students who were identified by school staff as not capable of participating meaningfully were excluded. The NAEP guidelines for exclusion were intended to assure uniformity of exclusion criteria from school to school as well as from state to state.

More information about the IEP/LEP questionnaire and exclusion criteria are provided in Chapter 5.

4.4.3 Teacher Questionnaires

To supplement the information on instruction reported by students, questionnaires were completed by the reading, U.S. history, or geography teachers of the fourth- and eighth-grade students who were assessed in those subject areas. The teacher questionnaires contained two parts. The first part pertained to the teachers' background and general training. The second part pertained to specific training in teaching the subjects being assessed, and the procedures the teacher used while giving instruction to assessed students.

Because the sampling for the teacher questionnaires was based on participating students, the responses to a particular teacher questionnaire do not necessarily represent all teachers of that subject area at that grade level in the nation. Rather, they are teachers of the representative sample of students assessed. It is important to note that in all NAEP reports, the student is always the unit of analysis, even when information from the teacher or school questionnaire is being reported. Using the student as the unit of analysis makes it possible to describe the instruction received by representative samples of students. Although this approach may provide a different perspective from other studies simply reporting information about teachers or schools, it is consistent with NAEP's goal of providing information about the educational context and performance of students.

The teacher questionnaires are described in detail in Chapter 2.

4.4.4 School Characteristics and Policies Questionnaire

The School Characteristics and Policies Questionnaire was given to the principal or administrator of each school that participated in NAEP. It included questions pertaining to

- background and characteristics of school principals,
- length of school day and year,
- school enrollment,
- absenteeism,
- drop-out rates,
- policies about tracking,
- curriculum,
- testing practices and use,
- special priorities and school-wide programs,
- availability of resources, special services, and community services,
- policies for parental involvement, and
- school-wide problems.

Chapter 5

FIELD OPERATIONS AND DATA COLLECTION[1]

Nancy W. Caldwell and Mark M. Waksberg
Westat, Inc.

5.1 INTRODUCTION

As a subcontractor to Educational Testing Service (ETS) on the National Assessment of Educational Progress (NAEP), Westat, Inc. was responsible for selecting schools and student samples, field operations, and data collection for the 1994 national component of NAEP. This chapter provides an overview of the field and data collection activities. A detailed discussion of these activities is given in the *Report on Data Collection Activities for the 1994 National Assessment of Educational Progress* (Westat, 1995).

The national assessment comprised main samples and long-term trend samples. In general, the main assessments involve new items and components; in the long-term trend assessments, elements of previous years' assessments are carried forward so that trends in student achievement can be measured over time. A total of 678 public and nonpublic schools took part in the long-term trend assessments of reading, mathematics, science, and writing, which were administered between October 11, 1993, and May 13, 1994. A total of 1,361 public and nonpublic schools took part in the main assessments of reading, U.S. history, and geography, which were administered between January 3 and April 1, 1994.

To reduce the burden on participating schools, NAEP field staff (hired and supervised by Westat) did most of the work associated with the assessments. NAEP staff contacted districts and schools in the fall to enlist cooperation, explain the assessment procedures, and to set a mutually agreed-upon assessment date. The NAEP supervisor visited the schools a week or two before the assessment to select the sample of students. The assessment sessions were conducted by the exercise administrators under the direction of the NAEP supervisors. At the completion of the assessment in a school, field staff coded the booklets and shipped the completed materials to National Computer Systems (NCS), the processing subcontractor.

5.1.1 Field Organization and Schedule of Activities

Three levels of staff—field managers, supervisors, and exercise administrators—were involved in the administration of the assessment. There were five field managers to whom the main assessment supervisors reported, and one field manager for the long-term trend assessment. Three of the main assessment field managers were in-house staff; the rest of the managers were field staff. Field managers monitored the activities of seven or eight supervisors, conducted quality control visits, and advised the field staff as appropriate. The field managers collected information from the supervisors on the status of the schools and the conduct of the assessments, and entered information on the number of students assessed into a computerized reporting system. The supervisors coordinated all aspects of the contacts with the schools and the assessments. They also supervised the exercise administrators, who conducted the actual assessments and helped with related clerical tasks.

---

[1] Mark M. Waksberg is responsible for survey operations and field activities for the NAEP assessments, under the direction of Nancy W. Caldwell.

Summarized below are the scheduled project activities for the 1994 main and long-term trend assessments.

Date	Activity
Mid-May 1993	Department of Education sent first letter to Chief State School Officers about the 1994 assessment. Westat sent state coordinators a list of their schools initially selected for either or both major components.
June-July 1993	ETS sent NAEP materials to districts to begin the process of obtaining cooperation of selected schools.
September 7-11 1993	Training sessions were held for long-term trend assessment supervisors and main assessment schedulers.
Early September - Mid-December 1993	Supervisors contacted district superintendents and nonpublic school principals about the assessment to establish or confirm participation with school representatives. Letter was sent to superintendents confirming the date, time, and place of the introductory meeting. A package of materials was sent to principals of participating schools. Westat selected substitutes for refusals.
October 11 - December 17, 1993	Fall long-term trend assessments were administered.
December 8-11, 1993	Main assessment supervisor training session was held.
January 3 - March 11, 1994	Winter long-term trend assessments were administered.
January 3 - April 1, 1994	Main assessments were administered.
March 14 - May 13, 1994	Spring long-term trend assessments were administered.

5.2 PREPARING FOR THE ASSESSMENTS

5.2.1 Gaining the Cooperation of Sampled Schools

The process of gaining cooperation of selected schools began in August 1993, with a series of letters and contacts with state and district officials. NCES and NAEP contacted the Chief State School Officer in each state notifying him or her of the districts and schools in the states that were in the sample. Throughout September and October 1993, Westat sent lists of sampled schools and other NAEP materials to district superintendents and heads of private schools, inviting their participation. These initial contacts paved the way for the telephone contacts by NAEP field supervisors to follow.

5.2.2 Supervisor Training

Training for assessment supervisors was multiphased and involved separate sessions conducted in September and December 1993. All training was conducted by the Westat project director, field director, and home office field supervisory staff. Also in attendance were representatives from ETS, NCS, and NCES. The first training session was held September 8-11, 1993, in Arlington, Virginia. Attending were the 25 supervisors who would work throughout the fall period to gain cooperation and schedule assessment activities with main NAEP schools, 5 troubleshooters, 13 supervisors responsible for conducting the long-term trend assessments, and the long-term trend field manager.

The session was split into two groups—one group was trained in the long-term trend assessment procedures and the other was trained in contacting the schools, gaining cooperation, and scheduling assessments. Other topics included supervisory responsibilities; a summary of various reports from recent assessments; recruiting and training exercise administrators; and administrative forms and procedures. The scheduling supervisors also received extensive training on using the reporting system installed on the laptop computers each of them was assigned.

Long-term trend supervisors and troubleshooters received specific instructions on procedures for drawing the sample of students, conducting assessments, and preparing and distributing assessment questionnaires. Also featured were practice exercises in sampling and in filling out the various administrative forms. A mock assessment session was held with the supervisors acting as "students." This included reading verbatim from an actual assessment script and following the prescribed procedures for distributing materials, reading directions, and recording the results of the assessment.

The 70 NAEP supervisors responsible for the main assessment were trained during the second session, held December 8-11, 1993. Training focused on a review of the preliminary activities during the fall (results of initial contacts with districts and schools, scheduling of assessments, and the status of exercise administrator recruitment), and assessment activities (sampling, excluding students, teacher surveys, and administrative forms and procedures).

The main assessment field managers were present to support training activities and answer questions concerning districts and schools that fell into the samples for both components of the assessment. Each supervisor also met with the person who completed the scheduling in his or her area, so that each school and district could be discussed in preparation for the new supervisors' contacts.

5.2.3 Recruiting, Hiring, and Training Exercise Administrators

During the fall, while the supervisors were contacting their schools and scheduling assessments, their other major responsibility was to recruit and hire exercise administrators, who would administer the assessment sessions. Exercise administrators were recruited from many sources. Each supervisor was given a PSU-by-PSU computerized list of administrators and interviewers who had worked for Westat previously. People who had served as exercise administrators before, with good evaluations from their previous supervisors, were usually the first considered for hiring. Subsequently, during contacts with the schools, the supervisors asked the school principals and other staff to recommend potential administrators. These referrals were frequently retired teachers or substitutes. Finally, where necessary, advertisements were placed in local newspapers and an employment service was notified.

Supervisors were told that, in general, three to five exercise administrators should be hired for each PSU, although a variety of factors might influence the actual number. The number of schools in a PSU, the size of the student sample in each school, distances to be traveled, the geography of the area, and weather conditions during the assessment period were all factors taken into consideration by supervisors in developing their plan for hiring administrators.

A few supervisors who had contiguous PSUs hired the same exercise administrators to work in all their PSUs. Supervisors who had PSUs with small and widely scattered schools tended to hire administrators to work only a portion of the PSU. Supervisors were encouraged to hire locally and to hire individuals with teaching experience or the ability to handle classroom situations.

The scheduling supervisors, all of whom were experienced NAEP supervisors, had complete responsibility for recruiting, hiring, and training all of the exercise administrators, including ones who would report to different assessment supervisors. In previous years, the supervisors had been given an outline of the training that they were to conduct for the administrators. In order to further standardize training, a script was prepared for training of administrators in 1994.

Each exercise administrator received a manual that covered the full range of their job responsibilities. After studying the manual, they attended a half-day training session. During the training, the supervisor reviewed, in detail, all aspects of the administrator's job including preparing materials, booklets, and Administration Schedules for assessments; the actual conduct of the session; post-assessment collection of materials; coding booklet covers; recordkeeping; and administrative matters.

5.2.4 Contacting Districts and Nonpublic Schools

After the supervisors were trained in September, they began working on obtaining cooperation. As supervisors contacted superintendents, principals, and nonpublic-school officials to introduce NAEP and determine the schools' cooperation status, they completed two forms and entered the school status in the receipt control system installed on their laptop computers. The Results of Contact Form was completed to document the discussion the supervisor had with each administrator concerning the district's willingness to participate and any special circumstances regarding the schools' cooperation or assessments.

The supervisor also completed portions of a School Control Form. This form was preprinted with the number and types of assessment sessions assigned to the school, so that information could then be shared with the district/school official. Information gathered during the phone call, including the name of the person designated to be the school coordinator, the number of students in the designated grade, tentative dates for the sampling visit and assessment, and other information that could have some bearing on the assessment, was recorded on the form. This information was used to update records in the home office. In December the forms were provided to the supervisors who would be conducting the assessments.

During the introductory presentation, the supervisor discussed arrangements for the assessments with representatives from each school. Within the weeks scheduled for the PSU, the supervisor had the flexibility to set each school's assessment date in coordination with school staff. The staff sometimes expressed preferences for a particular day or dates or had particular times when the assessment could not be scheduled. Using this information, the supervisors set up the assessment schedule for the PSU.

The supervisor usually learned during the introductory call or meeting whether a school required some form of parental notification or permission. Three versions of standard NAEP letters were prepared for schools' use. The first version informs parents about the assessment. The second assumes parental consent unless parents send the form back stating that they do not want their child to participate in the assessment. The third version requires that parents sign and return the form before students can be assessed. All versions of the letter were available to the schools, although when the issue of parental permission came up in discussion, supervisors offered the least restrictive version first. Schools could also send out their own letters and notices if they preferred not to use the ones prepared by NAEP. In addition, Spanish language versions of the informational parent letter were also made available to the schools. If a school required parents' letters, this information was recorded on the School Control Form.

After the completion of the introductory call or meeting, the completed School Control Forms were sent to the home office.

5.3 DATA COLLECTION FOR THE MAIN ASSESSMENTS

5.3.1 Selecting the Student Sample

After securing cooperation from the school, the first scheduled visit to be made to the school was to select the sample of students to take part in the assessment, and to conclude the arrangements for the actual testing. This visit was made in January by the supervisor responsible for the assessments in the school. Upon arriving at the school (rarely, sampling was done at the district office), the supervisor first reviewed the lists of age- and grade-eligible students. He or she confirmed with the school coordinator that all eligible students were listed. Age-eligible students were sometimes omitted because most standardized testing is done only with grade-eligible students. If any eligible students were omitted, sampling could not proceed until the list was completed.

Using the computer-generated Session Assignment Form, which was specific to the school, the supervisor selected the sample of students to be assessed. The form shows the type of sessions to be administered, the anticipated number of students to be assessed, the expected range of students eligible for the assessment, and then lists a series of line numbers. Those students who are listed on the line numbers shown on the form are those that are selected for the assessment. After making sure that all eligible students had been listed, the supervisor numbered the students. If the total number of eligible students was within the minimum and maximum limits indicated on the form, the supervisor proceeded to select the sample. If the number was outside the limits, he or she called Westat for additional sampling instructions before selecting the sample of students. The Session Assignment Forms provided step-by-step instructions for sampling, indicating not just the line number of each student to be selected, but the type of assessment for which each student was selected.

Once students were assigned to sessions, the supervisor and exercise administrators filled out an Administration Schedule for each session. The Administration Schedule is the primary control document for the assessment. It is used to list each sampled student and is the only link between booklets and students. Wherever possible, the supervisors tried to structure the assessment schedule so that each session would have about 30 or fewer students. The supervisor discussed the final schedule of the sessions with the school coordinator and the date, time, and location of each session were filled in on the Administration Schedules. Because student names are recorded on the Administration Schedules, those forms remained in the schools after the sample was drawn.

The supervisor then asked the coordinator to identify any students in the sample with an Individualized Education Plan (IEP) (for reasons other than being gifted and talented) and/or who were designated as Limited English Proficient (LEP). Any student with either (or both) of these designations was to be indicated on the Administration Schedules. The school was asked to complete an IEP/LEP student questionnaire for each student with this designation. (This was a change from earlier assessments, in which these questionnaires, then called excluded student questionnaires, were completed only for students who were actually excluded from the assessment.) The questionnaire was completed by a teacher, counselor, or other school official who knew the designated child well.

Next, the coordinator determined whether any of these students should be excluded from the assessment based on the criteria for excluding students (Figure 5-1).[2] About half of the students with an IEP or LEP were excluded from the assessment. If the school coordinator could not identify the excluded students while the supervisor was at the school, the instructions were left with the coordinator along with blank copies of the IEP/LEP student questionnaire. In those cases, the coordinator consulted with other school officials and informed the supervisor as to who was to be excluded when he or she returned for the assessment.

If requested by the school, the supervisor and/or exercise administrators made lists of the sampled students for their teachers and/or completed appointment cards notifying students about their assessment schedule. Teacher notification letters were also prepared in some schools that explained the assessment and listed the students that had been selected.

---

[2] Criteria differ for main and long-term trend assessments. The criteria for long-term trend assessments are given in Figure 5-2.

Figure 5-1

Instructions for Excluding Students (Main Assessments)

The intent is to assess all selected students. Therefore, all selected students who are capable of participating in the assessment should be assessed.

Some of the students identified on the Administration Schedule as Limited English Proficient (LEP) or as having an Individualized Education Plan (IEP) may be incapable of participating meaningfully in the assessment. School staff members knowledgeable about the IEP/LEP students, may exclude such students, as described below.

1. A student identified on the Administration Schedule as LEP may be excluded from the assessment if he/she:

   Is a native speaker of a language other than English;

   AND

   Has been enrolled in an English-speaking school (not including a bilingual education program) for less than two years;

   AND

   Is judged to be incapable of taking part in the assessment.

2. A student identified on the Administration Schedule as having an IEP or equivalent classification may be excluded from the assessment if:

   The student is mainstreamed less than 50 percent of the time in academic subjects and is judged incapable of participating meaningfully in the assessment;

   OR

   The IEP team or equivalent group has determined that the student is incapable of participating meaningfully in the assessment.

In some cases, IEP/LEP students meeting the above criteria should be assessed if, in the judgment of school staff, they are capable of taking the assessment.

WHEN THERE IS DOUBT, INCLUDE THE STUDENT

BEST COPY AVAILABLE

5.3.2 Conducting the Assessment Sessions

The primary responsibility for conducting assessment sessions was with the exercise administrators. Supervisors were required to observe the first session each administrator conducted to ensure that he or she followed the procedures properly. Supervisors were also required to be present in all schools with more than one small session to be conducted.

To ensure that sessions were administered in a uniform way, the exercise administrator was provided with scripts for each session type. The scripts were to be read verbatim. The scripts began with a brief introduction to the study. The administrator was then directed to distribute the booklets, being careful to match the student with the preassigned booklet.

After the booklets were distributed, some additional, scripted directions were read. Students were asked to write in the NAEP school ID and their home ZIP code on the cover of the booklet, and given some general directions in completing the assessment. For fourth-grade students, all of the background questions were then read aloud by the exercise administrator; at the older grades, the first question, which asks the students' race/ethnicity, was read by the administrator and the students read the rest to themselves. After the background questions were completed, the students were told that any further questions they might have could not be answered by the administrator, and were told to begin the first cognitive section of the assessment.

During the sessions, the exercise administrator walked around the room monitoring the students to make sure they were working in the correct section of their booklet and to prevent them from looking at a neighbor's booklet, although few assessed students in a given classroom were administered the same booklet.

At the end of an assessment session, booklets were collected and students were dismissed according to the school's policy. The exercise administrator was then responsible for completing the information at the top of the Administration Schedule, totaling the number of participating students, and coding the covers of all booklets, including those booklets assigned to absent students.

5.3.3 Participation Results for the Main NAEP Assessments

➪ School and Student Participation

The unweighted school response rate for the main assessments in 1994 was 83.4 percent. This is calculated by adding the cooperating and out-of-scope schools, and dividing that by the sum of the eligible schools and the refusals and out-of-scope schools. The final sample of cooperating schools included 399 schools at age 9/grade 4, 539 schools at age 13/grade 8, and 423 schools at age 17/grade 12. Table 5-1 shows comparative response rates for the last three assessments.

Table 5-1

Comparison of Student and School Response Rates (Unweighted) for Main Assessments, 1990-1994

Sample	1990	1992	1994
Student Response			
Age 9/Grade 4	92.9%	93.4%	93.2%
Age 13/Grade 8	89.0%	88.8%	91.0%
Age 17/Grade 12	80.8%	80.8%	81.1%
School Response			
Age 9/Grade 4	88.3%	86.4%	86.0%
Age 13/Grade 8	86.7%	85.3%	85.5%
Age 17/Grade 12	81.3%	81.5%	78.6%

Overall, 104,764 students were assessed across all three age/grade groups: 24,090 at age 9/grade 4, 41,728 at age 13/grade 8; and 38,946 at age 17/grade 12. The student participation rate across all age classes was 87.5 percent.

The response rate that would require a supervisor to conduct a makeup session was raised from the standard that had been used in previous main assessments. The old rate of 75 percent was changed to 85 percent—that is, any session (or group of sessions within the same subject area) at which less than 85 percent of the eligible students were assessed would require a makeup session (presuming that the school was willing to allow one). This change resulted in 117 more schools conducting makeup sessions than would have been previously required. In these schools, an additional 1,738 students were assessed, increasing the overall response rate by 1.7 percent, with the most significant increase occurring at age 17/grade 12.

➪ Assessment Questionnaires

Westat sent each school a school characteristics and policies questionnaire a few weeks before the assessment was scheduled to be conducted. Similarly, supervisors prepared an IEP/LEP student questionnaire for each sampled student with either of these designations, with the request that it be completed by someone at the school knowledgeable about that student.

Selected teachers of fourth- and eighth-grade reading, U.S. history, or geography were asked to fill out teacher questionnaires. The teachers asked to participate were the reading, U.S. history, or geography teachers of the students who were selected for assessment in those subject areas. The teacher questionnaire for grade 4 was combined into one form, since it is recognized that at this grade level, the same teacher would probably teach all of the subjects. For grade 8, there were two distinct questionnaires, one for reading and the other for teachers of U.S. history and/or geography.

The supervisor requested that the teacher questionnaires be distributed as quickly as possible after the sampling so that they could be returned by the day of the assessment. Additional introductory materials were included with the teacher questionnaires, in response to questions that teachers had about the importance of completing the questionnaires and about NAEP in general. Teachers received a letter explaining the purpose of the questionnaire, along with background materials about NAEP. In some cases, teachers received an additional letter which listed the names of their students who were selected for the assessments, if the school wished to notify the teachers in this way. If the teacher questionnaires were not complete at the time of the assessment, the supervisor left a postage-paid envelope to be used to return the questionnaires.

Table 5-2 shows the number of school, IEP/LEP, and teacher questionnaires distributed and the number and percentage of those that were returned.

Table 5-2
Response Data for Main Assessment Questionnaires

Age/ Grade	School Characteristics and Policies Questionnaires		IEP/LEP Student Questionnaires		Teacher Questionnaires			
					Reading		U.S. History/Geography	
	Expected	Returned	Expected	Returned	Expected	Returned	Expected	Returned
9/4	397	385 (97%)	3688	3359 (91%)	2030	1983 (98%)	—	—
13/8	541	553 (95%)	5685	5394 (95%)	2105	1938 (92%)	1645	1488 (91%)
17/12	417	439 (95%)	5135	4786 (93%)	—	—	—	—

5.3.4 The U.S. History Special Study

In an effort to find more instructionally relevant modes of assessment, NAEP developed an assessment of students' ability to work in groups to solve problems in the subject area of U.S. history. The purpose of this assessment was twofold: first, to explore the relationship between group communication behavior and group performance on written tasks pertaining to U.S. history; and second, to provide NAEP with experience in the design, administration, scoring and analysis of group tasks.

The U.S. history special study was conducted between April 18 and May 4, 1994, using a self-selected sample of 36 eighth-grade schools in which at least one U.S. history/geography session had been conducted as part of the 1994 main assessment. In each cooperating school, two group assessments sessions were scheduled: one session, entitled "1890s School" required the participation of four students; the other session, "Immigrants," required six students. Each assessment was videotaped and signed parental consent was sought for each student to be assessed.

➪ Field Organization

The field staff for the study was organized into six two-person teams, plus one team that acted as a backup. The primary role of the team leader was to secure school cooperation and to assume responsibility for the conduct of the assessment in their assignment of schools. The other team member was primarily responsible for operating the video equipment and working with the team leader to ensure a successful assessment. Both team members were required to complete observation protocols for each session administered in their assignment.

Supervisor training was held April 11-12, 1994, at ETS in Princeton, New Jersey. Training was conducted by the Westat project director and ETS project staff. The first day of training included an overview of the study and a description of the projects; a simulated demonstration of the administration of both session tasks; an overview of the group communication ratings; and practice sessions for recording each of the three categories of communications ratings followed by a "live" administration of the two projects using age-appropriate participants.

The second day of training focused on a review of the rating scales. In order to determine interrater reliability among the supervisors, group communication protocols were completed for an assessment previously videotaped by ETS in one of the pilot schools. The completed protocols were collected and scored by ETS staff; the interrater reliability was determined to be 83 percent. A videotape of pilot sessions conducted by ETS was distributed to supervisors to provide additional practice in completing communication ratings and to increase interrater reliability.

Supervisors also received instruction on administration procedures such as sampling, preparing session document forms, arranging the assessment room, handling requests for viewing assessment materials, and responding to students' questions. Instruction was provided on the use of the video equipment as well as completing post assessment activities such as distributing the Student Reaction Sheet, completing administrative forms, and packing and shipping assessment materials back to ETS.

➪ School Sample

The school sample consisted of 36 eighth-grade schools where at least one U.S. history/geography session had been conducted as part of the 1994 main assessment. Once interested schools had been identified, a sample of 36 eligible schools in 17 PSUs was selected in order to represent diversity in school size, school type, minority population, and geographical area.

After a school agreed to cooperate, the supervisor confirmed with the school coordinator that a complete copy of the Administration Schedule left at the school at the conclusion of the main NAEP assessment had been retained and could be located. The Administration Schedule was then used to determine which students were eligible to be assessed for the study.

⇨ Student Sample

The selection of the student sample was two-fold; the first step required identifying students eligible for the study, according to the criteria described in the following section. Then, from those students identified as eligible, a random sample was drawn from those for whom signed parental consent was obtained and who were present on assessment day.

After an assessment date and time had been established, the supervisor provided the school coordinator with instructions on how to identify eligible students from those listed on the Administration Schedule. Students were eligible if they were in the eighth grade and had been assessed in U.S. history.

Once eligible students had been identified, a videotape release form was distributed to the parents/guardians of each eligible student by the school coordinator to be returned no later than assessment day. It was at the discretion of the school coordinator whether to mail release forms home to parents or send them home with students. If requested, postage fees were paid by NAEP if the consent forms were to be mailed.

Because the confidentiality of participants in the U.S. history special study was guaranteed by the same authorizing statute as that of the rest of NAEP, a statement of guidelines concerning confidentiality was provided to schools. The guidelines included a requirement that students could not be assessed without a signed videotape release form, which was left at the school after the study with student identification obliterated. The resulting videotapes were also held confidential.

For each signed consent form, the student's eligibility was confirmed, using the information recorded on the Administration Schedule(s). If enough students met the eligibility criteria, a random sample of ten students was selected, four randomly assigned to the "1890s School" project and six to the "Immigrants" session. If there were not enough students to assign to both sessions, a requisite sample was randomly drawn from those eligible and assigned to whichever session could be administered with the number of students who met the criteria.

⇨ Administering the Study

Each session was administered using a standardized session script and took approximately one hour to complete. One team member, usually the team leader, was responsible for conducting the session; the other was responsible for videotaping the session. A tape recorder was provided to each team in the event that the videocamera malfunctioned and an audio backup was required.

The link to the 1994 NAEP assessment as well as student confidentiality was preserved by having each team complete a Session Summary Sheet. The student's NAEP U.S. history booklet ID number (from the 1994 assessment) was recorded on one side of the sheet next to a student number that was associated with each booklet ID number. The student number was used to illustrate the seating arrangement at the bottom of the sheet to assist with scoring. Student names, recorded to the left of the booklet ID, were detached at the conclusion of the assessment and were left at the school.

Following the conclusion of the session, students were asked to complete a Student Reaction Sheet in order to obtain information about curriculum instruction they may have received relating to the topic assessed, to solicit students' evaluation of their input in response to the task, and their overall reaction to the session.

A total of 35 schools cooperated in the study. In the schools sampled for this study, approximately 750 students were assessed in U.S. history in the 1994 NAEP. As noted earlier, the original plan was to assess a total of ten students in two different groups in each school; that is, a target of 350 students. However, in half of the participating schools only one U.S. history special study session was able to be conducted, and only 257 students were assessed (73 percent of the target).

There were several factors that contributed to this lowered participation. First, because of time and cost constraints, the responsibility of identifying eligible students for the study and distributing the Videotape Release Form to parents was assigned to the school coordinator. The task of identifying eligible students was not entirely straightforward since, in most schools, there were three different session types (reading, U.S. history and geography). Quality control measures were difficult to impose until the day of the assessment when the project team was at the school. It was found that in some schools the coordinator identified ineligible students because they used the Administration Schedule for the wrong session. In other cases the information included on the Administration Schedule was not complete. Consequently, not all eligible students were sampled and therefore Videotape Release Forms were not distributed to parents of all eligible students, further limiting the pool of students eligible to participate. Overall, (37 percent) of the consent forms distributed were returned. Only eight refusals were received.

Another factor restricting response was that the session required exactly six and exactly four students. If, for example, there were only nine eligible students for whom permission had been obtained then only six or only four of these students could be assessed.

5.4 DATA COLLECTION FOR THE LONG-TERM TREND ASSESSMENTS

To provide continuity and comparability with past NAEP studies, the long-term trend assessments repeat procedures and materials that have been used in previous assessments. The fall trend assessment of age 13/grade 8 students was held in the ten-week period from October 11-December 17, 1993; the winter assessment of age 9/grade 4 students was held during the ten-week period from January 3 through March 11, 1994; and the spring trend assessment of age 17/grade 11 students in the nine-week period from March 14 through May 13, 1994. Students were assessed in reading, writing, mathematics, and science.

Audiotape-administered sessions were conducted with samples of age-eligible students only, as had been done in all previous years. Additional samples of age- and grade-eligible students were assessed with print-administered booklets, following procedures initiated in the 1984 assessment. Six different types of sessions were conducted—one print-administered session and five different tape-administered sessions. Depending on the size of the participating school, up to three different session types, involving a total of 80 students, were conducted there.

5.4.1 Selecting the Student Sample

Procedures for sampling in the long-term trend schools were very similar to those used in the schools selected for the main assessment. One to two weeks before the assessment, the supervisor went to the schools to make this selection. Lists of students were reviewed to ensure that all age- and grade-eligible students were listed. The Session Assignment Form specified a range between which the number of students should fall. As long as the total number of students was within this allowable range, the sampling proceeded. Students were assigned to a session type based on their line number from the list that the school had prepared.

A significant difference in the sampling for the long-term trend assessments was that for the tape-administered sessions, only age-eligible students were selected. For those sessions, the supervisor selected from the entire list of students (age- and grade-eligible), but then deleted those who were only grade-eligible before recording the names of the students to be assessed on the Administration Schedules.

The criteria for excluding students were also different for the long-term trend schools, again, employing the criteria that were established previously. For those students who were excluded, the school was asked to complete an excluded student questionnaire. If the coordinator could not identify the excluded students while the supervisor was at the school, a set of instructions for excluding students (Figure 5-2) was left with the coordinator along with the estimated number of questionnaires needed.

5.4.2 Conduct of the Assessment

The conduct of the assessments in schools selected for the long-term trend assessments was essentially the same as in the main schools. Scripts were provided for the supervisors and exercise administrators to administer the sessions. The major difference was in the administration of the tape sessions. In these sessions, after the distribution of the test booklets, the administrator turned on the tape recorder. The remainder of the instructions were announced from the tape, and the timing of the assessment was paced by the length of time that the tape ran.

5.4.3 Participation Results for the Long-Term Trend Assessments

➪ School and Student Participation

The unweighted school response rate for the long-term trend assessments was 83.2 percent. The final sample of cooperating schools included 243 schools at age 9/grade 4, 240 schools at age 13/grade 8, and 195 schools at age 17/grade 11. The student participation rate across all age classes was 90.5 percent.

Of the 36,912 students sampled for assessments, 6.4 percent were excluded by schools. Overall, 31,250 students were assessed across all three age/grade groups: 10,998 students were assessed at age 9/grade 4, 11,599 students were assessed at age 13/grade 8, and 8,653 were assessed at age 17/grade 11. Table 5-3 shows comparative response rates for the last three trend assessments.

Figure 5-2

Instructions for Excluding Students (Long-Term Trend Assessments)

These instructions accompany the NAEP Administration Schedules listing the students selected to participate in the National Assessment. Please review the Administration Schedule and, in concert with other school officials, determine whether any students should be excluded from the assessment because they are non-English speaking, educable mentally retarded, or functionally disabled. These categories are defined as follows:

- **Non-English speaking students** - Those who do not read or speak English and would be unable to overcome the language barrier in the test situation.

- **Educable mentally retarded (EMR)** - Students who have been psychologically tested as EMR students or students who are considered EMR in the professional opinion of the principal or other qualified staff members. However, students should not be excluded because of poor academic performance or normal discipline problems. Only those students should be excluded who cannot give meaningful responses to exercises at their age level.

- **Functionally disabled** - (temporary or permanent physical disability) Students who are so disabled that they cannot perform in the NAEP testing situation should also be excluded. However, functionally disabled students who can respond should be included.

If the school determines that a student should be excluded, the student's name should be lined through, taking care to make sure that the name is still legible.

Next, prepare an Excluded Student Questionnaire for each sampled student the school has determined should be excluded from the assessment. Write the student's name in the area labeled "Teacher of -" and give the questionnaire to the staff member most knowledgeable about the student.

The Excluded Student Questionnaire should be returned by the day of the assessment and given to the NAEP Supervisor or Exercise Administrator who will complete the coding on the cover of the questionnaire.

Table 5-3

Comparison of Student and School Response Rates (Unweighted) for Long-Term Trend Assessments, 1990-1994

Sample	1990	1992	1994
Student Response			
Age 9/Grade 4	92.4%	94.0%	94.2%
Age 13/Grade 8	90.4%	90.8%	92.2%
Age 17/Grade 11	81.2%	82.8%	84.1%
School Response			
Age 9/Grade 4	88.1%	87.4%	86.7%
Age 13/Grade 8	90.5%	84.7%	81.7%
Age 17/Grade 11	80.7%	81.3%	81.1%

➪ Assessment Questionnaires

The school characteristics and policies questionnaire and the excluded student questionnaire were distributed in the schools to be completed by school personnel.

The school characteristics and policies questionnaire was brought to the school by the assessment supervisor at the time of the sampling visit. This form was to be filled out by the principal or other staff member knowledgeable about the school's administrative policies and staff characteristics. The supervisors collected the completed questionnaire when they returned to the school for the assessment.

An excluded student questionnaire was filled out for every student who was sampled for the assessment but excluded by the school. Following exclusion criteria used in previous trend assessments, schools could exclude students with limited English-speaking ability, those who were educable mentally retarded, or functionally disabled students, if in the judgment of school staff or if school records indicated they were unable to "participate meaningfully" in the assessment. After the sample of students was drawn and Administration Schedules prepared, the supervisor requested that the school coordinator identify any students who should be excluded. The supervisor then gave an excluded student questionnaire to the coordinator for every excluded student, with the request that it be completed by someone in the school knowledgeable about the student's disability.

The supervisor attempted to collect all completed questionnaires on the assessment day. If the questionnaires were not ready, and it was convenient for the supervisor or an exercise administrator to return to the school later to pick them up, they would do so. Otherwise, the supervisor gave the coordinator a postage-paid envelope to use to mail the forms to NCS. Of the 675 school characteristics and policies questionnaires distributed to the schools, 652 were completed and collected by NAEP supervisors. This represents a completion rate of 96.6 percent. The return of excluded student questionnaires was comparably high, at 98.3 percent; 2,452 of 2,494 questionnaires were returned.

Once the assessments were completed in a school, the supervisor and exercise administrator completed the coding of the front covers of the assessment booklets, filled out the necessary forms, and shipped the booklets and forms to NCS. A copy of all forms was sent to Westat so that progress in the field could be closely monitored.

The School Worksheet was used by the supervisor to summarize the results of the assessment sessions in each school. The number of students to be assessed, the number actually assessed, and those absent were entered so that the supervisor could determine whether a makeup session was required. Attendance of less than 86 percent required a makeup. If a makeup was required for one or more session types, the supervisor discussed the scheduling of the makeup with the coordinator.

Previously in long-term trend assessments, the percentage of students attending that would necessitate a makeup session was 75 percent or below. By raising that rate to 85 percent, an additional 236 students were assessed in 49 schools, increasing the overall response rate by less than one percent.

The top (original) copy of the School Worksheet, any excluded student questionnaires completed by the school, and the Administration Schedules (with the students' names removed and left at the school) were included with the booklets in the shipment to NCS. In addition, the supervisor included a Packing List with the materials, which inventoried the assessment materials sent to and returned from the school.

5.5 FIELD MANAGEMENT

Two field managers assisted Westat home office staff in monitoring the field work. Some of the assessment supervisors responsible for main NAEP assessments reported to these field managers, who in turn reported to the Westat field director. The remaining main NAEP supervisors reported directly to three members of Westat's home office staff. Supervisors and field managers were required to report to their contacts at least once a week.

An automated management system was developed and maintained in Westat's home office. The scheduling supervisors working in the fall had this system installed on the laptops that they used. The system contained a record for each sampled school. A disposition code structure was developed to indicate the status of each school's participation (e.g., school cooperating, decision pending, school refusal, district refusal, school closed, etc.). As a school's status was determined, the supervisors would enter the status of the school onto their laptops, and this information was downloaded onto the home office system on a weekly basis. Disposition reports were then generated from the receipt system once a week in order that home office staff could review the progress of securing cooperation from the sampled schools.

These reports were an invaluable tool for the sampling statisticians as well as for the field director and field management staff. They provided the statisticians with the information needed to determine whether the sample of schools was adequate to produce representative results. Based on the information contained in these reports, the sampling statisticians selected substitute schools to replace some of the noncooperating schools.

After assessments were completed, data from the School Worksheets on the number of students to be assessed, the number assessed, and the number absent were entered for each school. Data on completed questionnaires received was provided by NCS. Reports were generated weekly that allowed the project staff to monitor the progress of the assessments both in terms of checking that the schools were assessed on schedule as well as assuring that a high response rate was achieved. The sampling statisticians used these reports to monitor the sample yield by school, PSU, and age/grade level.

Progress of the assessments was constantly monitored through telephone reports held between NAEP supervisors, field managers, and home office staff. During these phone conversations, the supervisors' schedules were reviewed, and any problems that the supervisors were experiencing were discussed. Much of the attention this year was focused on maintaining the schedule in light of the many postponements and cancellations due to the winter weather and earthquake.

The supervisors who traveled filled out a Work Schedule for a one- to two-week period, showing their whereabouts, so that they could be contacted if necessary. It also allowed field managers and project staff to review the supervisors' schedules and the distribution of work.

Progress of the field work was also monitored during quality control visits made to the field by Westat and ETS office staff.

Chapter 6

PROCESSING ASSESSMENT MATERIALS[1]

Patrick B. Bourgeacq and Timothy Robinson
National Computer Systems

6.1 INTRODUCTION

This portion of the report reviews the activities conducted by National Computer Systems (NCS) for the NAEP 1994 main assessments in reading, U.S. history, and geography, and long-term trend assessments in reading, mathematics, science, and writing. NCS is a global company providing data collection services and systems to selected segments of the education, business, government, and healthcare markets. NCS's Measurement Services Division (MSD) is the major service bureau of NCS, providing large-scale data processing and test processing services to customers nationwide, including state education agencies, major test publishers, and the U.S. Department of Education, as well as numerous other federal agencies and commercial organizations. As a subcontractor to Educational Testing Service (ETS), NCS was responsible for printing all of the NAEP student booklets and control documents; distributing the assessment materials to the field; receiving, tracking, processing and editing the assessment materials as they returned from the field; scoring all of the constructed-response items (in conjunction with ETS); and delivering the assessment data files to ETS for analysis and reporting.

The 1994 assessment was an exciting one for NAEP and NCS because of the introduction of image scoring to the assessment. The advent of image scoring eliminated almost all paper handling during scoring and improved monitoring and reliability scoring. A short-term trend study was added to the assessment to compare the scoring of paper and scoring of images of student responses from both 1992 and 1994.

In the early 1990s, NCS developed and implemented flexible, innovatively designed processing programs and a sophisticated process control system that allowed the integration of data entry and work flow management systems. The planning, preparation, and quality-conscious application of these systems in 1992 and 1994 has made the NAEP project an exercise in coordinated teamwork and excellence.

This chapter begins with a description of the various tasks performed by NCS, detailing printing, distribution, receipt control, scoring, and processing activities. It also discusses specific activities involved in processing the assessment materials, and presents an analysis of several of those activities. The chapter provides documentation for the professional scoring effort—scoring guides, training papers, papers illustrating sample score points, calibration papers, calibration bridges, and interreader reliability reports. The detailed processing specifications and documentation of the NAEP process control system are presented later.

---

[1] Patrick B. Bourgeacq is the NCS/NAEP project director for scoring. Timothy Robinson is the senior coordinator of processing operations for the project.

6.1.1 Innovations for 1994

Much of the information necessary for documentation of accurate sampling and for calculating sampling weights is collected on the administration schedules which, until 1993, were painstakingly filled out by hand by Westat administrative personnel. In 1994, for the first time, much of the work was computerized—booklets were preassigned and booklet ID numbers were preprinted on the administration schedule. When Westat personnel received the documents, they filled in only the "exception" information. This new method also permitted computerized updating of information when the administration schedules were received at NCS, eliminating the need to sort and track thousands of pieces of paper through the processing stream.

The introduction of image processing and image scoring further enhanced the work of NAEP. In image processing, answer documents are scanned and the actual images of the student responses are transferred directly to data files. (The various aspects of image processing are explained in detail later in this chapter.) The scanned images are then routed electronically to reader workstations. During the image scoring phase, the reader will access the images and enter the scores directly into the database. This allows item-by-item scoring rather than the traditional scoring by block, thereby improving accuracy and speed of scoring. Imaging technology allows data and images to be effectively managed and securely maintained in digitized data form rather than in hard copy format.

Image processing and scoring were successfully piloted in a side-by-side study conducted during the 1993 NAEP field test, and so became the primary processing and scoring methods for the 1994 assessment. Image processing allowed the automatic collection of handwritten demographic data from the administrative schedules and the student test booklet covers through intelligent character recognition (ICR). This service was a benefit to the jurisdictions participating in NAEP because they were able to write rather than grid certain information—a significant reduction of burden on the schools. Image processing also made image scoring possible, eliminating much of the time spent moving paper. The images of student responses to be scored were transmitted electronically to the scoring center, located at a separate facility from where the materials were processed.

The success of this new method of transferring data has moved NAEP closer to achieving another goal—the simultaneous scoring of constructed-response items at multiple locations. This process enhanced the reliability and monitoring of scoring and allowed both NCS and ETS to focus attention on the intellectual process of scoring student responses.

Tables 6-1 and 6-2 give an overview of the processing volume and the schedule for the 1994 NAEP assessment.

Table 6-1

Processing Totals for the 1994 NAEP Assessment

Document/Category	Totals
Number of sessions	6,521
Assessed student booklets	105,538
Absent student booklets	18,094
Excluded student questionnaires	2,452
IEP/LEP questionnaires	13,539
School questionnaires	1,943
Teacher questionnaires	5,364
Scanned documents	136,960
Scanned sheets	3,270,232
Key-entered documents	18,983

Table 6-2

1994 NAEP Assessment Schedule of Printing and Processing Activities

Phase	Count	Activity	Plan Start Date	Plan Finish Date	Actual Start Date	Actual Finish Date
Fall Long-Term Trend	29,000	Printing	7/1/93	8/29/93	7/1/93	9/1/93
		Initial packaging begins	8/23/93	9/3/93	8/23/93	9/3/93
		Session data file to NCS from Westat	8/30/93	8/30/93	8/30/93	8/30/93
		Print administration schedule	9/3/93	9/3/93	9/3/93	9/3/93
		Last date all materials to NCS	9/3/93	9/3/93	9/10/93	9/10/93
		Final packaging	9/3/93	9/17/93	9/3/93	9/17/93
		Ship administration schedule	9/7/93	9/7/93	9/7/93	9/7/93
		Westat supervisor training-trend assessments	9/8/93	9/11/93	9/8/93	9/11/93
		Final address file to NCS from Westat	9/9/93	9/9/93	9/9/93	9/9/93
	25,600	Distribution	9/20/93	10/4/93	9/20/93	10/4/93
		Materials due in districts	9/24/93	9/24/93	9/24/93	9/24/93
		Test administration	10/11/93	12/17/93	1/15/94	1/15/94
		Receiving	10/13/93	12/21/93	1/15/94	1/15/94
		Scoring training preparation	10/18/93	10/29/93	10/18/93	10/29/93

BEST COPY AVAILABLE

Table 6-2 (continued)

1994 NAEP Assessment Schedule of Printing and Processing Activities

Phase	Count	Activity	Plan Start Date	Plan Finish Date	Actual Start Date	Actual Finish Date
	11,200	Processing	10/18/93	12/23/93	10/18/93	1/15/94
		Scoring training	11/3/93	11/12/93	11/2/93	11/5/93
		Constructed-response scoring	11/15/93	1/21/94	11/2/93	1/18/94
		Tape delivered	1/28/94	1/28/94	1/28/94	1/28/94
Main		Printing meeting NCS/Columbia	3/8/93	3/8/93	3/8/93	3/8/93
		Meeting with NCES	5/17/93	5/18/93	5/17/93	5/18/93
		NCS receives short-term trend boards from ETS	6/1/93	6/1/93	6/7/93	6/1/93
		Review by Instrument Development Committee	6/1/93	6/30/93	5/12/93	5/28/93
		Network meeting to review reading items in DC	6/18/93	6/18/93	6/18/93	6/18/93
		Need information on assignment of NAEP Readers	6/30/93	6/30/93	6/30/93	6/30/93
		Image development–archive system available	7/6/93	7/15/93	7/6/93	7/15/93
		Image meeting at NCS	7/7/93	7/8/93	7/7/93	7/8/93
		NCS receives camera-ready copy for reading	7/25/93	8/5/93	8/10/93	8/25/93
		Image development–hardware vendor selection done	7/30/93	7/30/93	7/30/93	7/30/93
		Administration schedule from printers	8/1/93	8/1/93	8/11/93	8/11/93
		NCS receives camera-ready copy for U.S. history	8/13/93	8/24/93	8/26/93	9/2/93
		NCS receives all background questions	8/15/93	9/5/93	8/20/93	8/27/93

Table 6-2 (continued)

1994 NAEP Assessment Schedule of Printing and Processing Activities

Phase	Count	Activity	Plan Start Date	Plan Finish Date	Actual Start Date	Actual Finish Date
Main (continued)		NCS receives camera-ready copy for geography	8/20/93	9/2/93	9/3/93	9/15/93
		Image development–remote site selected	8/24/93	8/24/93	8/24/93	8/24/93
		Image development–hardware plan complete	8/31/93	8/31/93	8/24/93	8/24/93
		ETS receives NAGB approval and OMB clearance for reading	9/1/93	9/1/93	8/25/93	8/25/93
	726,000	Printing	9/2/93	12/1/93	9/2/93	11/29/93
		Image development–trend code complete	9/17/93	9/17/93	9/19/93	9/19/93
		ETS receives NAGB approval and OMB clearance for geography	9/24/93	9/24/93	9/14/93	9/14/93
		ETS receives NAGB approval and OMB clearance for U.S. history	9/30/93	9/30/93	9/2/93	9/2/93
		NCS submits receipt-control specifications plan	10/1/93	10/1/93	10/1/93	10/1/93
		All reading booklets at NCS	10/15/93	10/15/93	11/29/93	11/29/93
		Initial packaging begins	10/15/93	12/20/93	10/12/93	12/18/93
		Weekly status reports on receipt-control and procedures	10/18/93	5/31/94	10/18/93	6/14/94
		Subcontractor meeting	11/8/93	11/9/93	11/8/93	11/9/93
		Image development–scanning definition	11/29/93	1/14/94	11/29/93	1/14/94
		All materials at NCS	12/1/93	12/1/93	12/15/93	12/15/93
		95% of session data to NCS from Westat	12/3/93	12/3/93	12/3/93	12/3/93
		Westat training for national supervisors	12/6/93	12/10/93	12/6/93	12/10/93
		Print administration schedule	12/8/93	12/8/93	12/8/93	12/8/93

BEST COPY AVAILABLE

Table 6-2 (continued)

1994 NAEP Assessment Schedule of Printing and Processing Activities

Phase	Count	Activity	Plan Start Date	Plan Finish Date	Actual Start Date	Actual Finish Date
Main (continued)		Ship administration schedule	12/10/93	12/10/93	12/10/93	12/10/93
		Final address file to NCS from Westat	12/17/93	12/17/93	12/17/93	12/17/93
		Address file for Main samples from Westat	12/17/93	12/17/93	12/17/93	12/17/93
		Final packaging	12/20/93	12/23/93	12/20/93	12/23/93
	330,000	Distribution–bulk and Wave I	12/27/93	12/27/93	12/27/93	12/27/93
		Materials due in districts	1/2/94	1/2/94	1/2/94	1/2/94
		Scoring training preparation for reading	1/3/94	3/25/94	1/3/94	3/25/94
		Scoring training preparation for U.S. history	1/3/94	3/25/94	1/3/94	3/25/94
		Scoring training preparation for geography	1/3/94	3/25/94	1/3/94	3/25/94
		Test administration	1/3/94	4/8/94	1/3/94	4/18/94
		Receiving	1/5/94	4/15/94	1/5/94	4/22/94
	198,100	Processing	1/7/94	5/6/94	1/7/94	5/20/94
		Image development–clip areas complete	1/14/94	1/14/94	1/14/94	1/14/94
		Image development–edits complete	1/14/94	1/14/94	1/14/94	1/14/94
		Image development–scanning begins	1/17/94	1/17/94	1/17/94	1/17/94
		Image development–remote site open	1/17/94	1/17/94	1/17/94	1/17/94
		Final packaging Wave 2 materials	1/21/94	1/21/94	1/21/94	1/21/94
		Ship Wave 2 materials	1/28/94	1/28/94	1/28/94	1/28/94
		Wave 2 materials due in districts	2/2/94	2/2/94	2/2/94	2/2/94
		Image development–scoring definitions complete	2/10/94	2/10/94	2/17/94	2/17/94
		Image development–overlays and OEs complete	2/10/94	2/10/94	2/17/94	2/17/94
		Image development–initial CAR distribution	2/11/94	2/11/94	2/20/94	2/20/94

Table 6-2 (continued)

1994 NAEP Assessment Schedule of Printing and Processing Activities

Phase	Count	Activity	Plan Start Date	Plan Finish Date	Actual Start Date	Actual Finish Date
Main (continued)		Image development–training selection complete	2/11/94	2/11/94	2/20/94	2/20/94
		Final packaging Wave 3	2/18/94	2/18/94	2/12/94	2/12/94
		Demo of scoring for ETS	2/21/94	2/21/94	3/3/94	3/3/94
		Ship Wave 3 materials	2/25/94	2/25/94	2/25/94	2/25/94
		Wave 3 materials due in districts	3/2/94	3/2/94	3/2/94	3/2/94
		Subcontractor meeting	3/28/94	3/29/94	3/28/94	3/28/94
		Constructed-response scoring/training for reading	3/28/94	5/13/94	3/28/94	6/10/94
		Constructed-response scoring/training for U.S. history	3/28/94	5/13/94	3/28/94	5/24/94
		Constructed-response scoring/training for geography	3/28/94	5/13/94	3/28/94	5/20/94
		Project through clean post	4/29/94	4/29/94	5/2/94	5/2/94
		Ship weights data tape to Westat	5/6/94	5/6/94	5/6/94	5/6/94
		Tape delivered for questionnaires	5/23/94	5/24/94	5/23/94	5/24/94
		Tape delivered for U.S. history	6/2/94	6/3/94	6/2/94	6/3/94
		Tape delivered for geography	6/9/94	6/10/94	6/9/94	6/10/94
		Tape delivered for reading Main/all grades	6/23/94	6/24/94	6/23/94	6/24/94
		Subcontractor meeting	7/11/94	7/12/94	7/11/94	7/12/94
Winter Long-Term Trend	29,000	Printing	7/1/93	10/31/93	7/1/93	10/31/93
		Initial packaging begins	10/25/93	12/10/93	10/25/93	12/10/93
		Session data file to NCS from Westat	11/22/93	11/22/93	11/22/93	11/22/93
		Print administration schedule	11/22/93	11/22/93	11/22/93	11/22/93
		Last date all materials to NCS	12/3/93	12/3/93	12/3/93	12/3/93
		Final address file to NCS from Westat	12/3/93	12/3/93	12/3/93	12/3/93
		Ship administration schedule	12/3/93	12/3/93	12/3/93	12/3/93
		Final packaging	12/10/93	12/15/93	12/10/93	12/15/93

Table 6-2 (continued)

1994 NAEP Assessment Schedule of Printing and Processing Activities

Phase	Count	Activity	Plan Start Date	Plan Finish Date	Actual Start Date	Actual Finish Date
Winter Long-Term Trend	25,600	Ship Wave 1 and bulk materials	12/17/93	12/17/93	12/17/93	12/17/93
		Wave 1 and bulk materials due in districts	12/22/93	12/22/93	12/22/93	12/22/93
		Final packaging Wave 2 materials	1/1/94	1/1/94	1/1/94	1/1/94
		Test administration	1/3/94	3/18/94	1/3/94	3/25/94
		Receiving	1/5/94	3/14/94	1/5/94	3/27/94
	11,200	Processing	1/5/94	6/17/94	1/5/94	6/17/94
		Ship Wave 2 materials	1/21/94	1/21/94	1/21/94	1/21/94
		Scoring training preparation	1/24/94	1/28/94	1/24/94	1/28/94
		Scoring training	1/31/94	2/4/94	1/31/94	2/4/94
		Constructed-response scoring	2/7/94	4/1/94	2/7/94	4/1/94
		Project through clean post	4/1/94	6/19/95	4/1/95	6/19/94
		Tape delivered	7/1/94	7/1/94	7/1/94	7/1/94
Spring Long-Term Trend	24,000	Printing	7/1/93	10/31/93	7/1/93	10/31/93
		Initial packaging begins	1/11/94	2/11/94	1/11/94	2/11/94
		Last date all materials to NCS	2/2/94	2/2/94	2/2/94	2/2/94
		Session data file to NCS from Westat	2/5/94	2/5/94	2/5/94	2/5/94
		Print administration schedule	2/11/94	2/11/94	2/11/94	2/11/94
		Final address file to NCS from Westat	2/12/94	2/12/94	2/12/94	2/12/94
		Final packaging	2/14/94	2/23/94	2/14/94	2/23/94
		Ship administration schedule	2/16/94	2/16/94	2/16/94	2/16/94
	22,200	Distribution: Wave I	2/21/94	2/21/94	2/21/94	2/21/94
		Materials due in districts: Wave I	2/25/94	2/25/94	2/25/94	2/25/94
		Test administration	3/14/94	5/13/94	3/14/94	5/13/94
		Receiving	3/16/94	5/20/94	3/16/94	5/20/94

Table 6-2 (continued)
1994 NAEP Assessment Schedule of Printing and Processing Activities

Phase	Count	Activity	Plan Start Date	Plan Finish Date	Actual Start Date	Actual Finish Date
Spring Long-Term Trend	9,200	Processing	3/18/94	6/17/94	3/20/95	6/24/94
		Distribution: Wave II	3/28/94	3/28/94	3/28/94	3/28/94
		Materials due in districts: Wave II	4/1/94	4/1/94	4/1/94	4/1/94
		Scoring training preparation	4/4/94	4/15/94	4/4/94	4/15/94
		Scoring training	4/16/94	4/16/94	4/16/94	4/16/94
		Constructed-response scoring	4/18/94	6/3/94	4/11/94	6/10/94
		Processing through clean post	5/27/94	5/27/94	6/3/94	6/3/94
		Tape delivered	7/1/94	7/1/94	7/1/94	7/1/94
Holistic Writing		Processing	8/12/94	8/12/94	8/12/94	8/12/94
		Scoring training preparation	6/5/94	6/22/94	7/6/94	7/8/94
		Training and constructed-response scoring	6/20/94	6/24/94	7/11/94	7/15/94
		Tape delivered	8/12/94	8/12/94	8/12/94	8/12/94
Mechanics of Writing		Processing	8/12/94	8/12/94	8/12/94	8/12/94
		Scoring training preparation	5/30/94	6/10/94	7/5/94	7/15/94
		Scoring training	7/5/94	7/9/94	7/18/94	7/20/94
		Constructed-response scoring	7/10/94	8/5/94	7/21/94	7/29/94
		Tape delivered	8/12/94	8/12/94	9/1/94	9/1/94

6.2 PRINTING

For the 1994 assessment, 242 discrete documents were designed. More than one million booklets and forms, totaling over 28 million pages, were printed. The printing effort began in June 1993, with the design of the booklet covers and the administration schedule. This was a collaborative effort involving staff from ETS, Westat, and NCS. The covers were designed to facilitate the use of intelligent character recognition to gather data. The administration schedule, which was designed to use both intelligent character and optical mark recognition, was designed to be the primary source of demographic data and to also serve as the session header for booklets when processed. Spaces for the same information were included on the student booklet cover as a backup source. For elements not individualized on the administration schedule (school number, ZIP code, ILSQ Number, and "do not use" field), both handwritten information and optical mark recognition ovals were used on the booklet cover to assure complete, accurate data collection.

Printing of the 1994 NAEP documents began with the materials needed for the long-term trend assessments. These included the Administration Schedule, the 26 assessment booklets, the excluded student questionnaire, the grade 8 school characteristics and policies questionnaire and the roster of school and excluded student questionnaires. All materials for the long-term trend assessments except the grade 8 school questionnaires were ready on or before September 1, 1993. The grade 8 school questionnaires followed on September 20, 1993. The printing of assessment booklets and questionnaires for the main assessments followed directly. Delivery of these documents was complete by November 29, 1993.

Details of the printing procedures are given in the *Report of Processing and Professional Scoring Activities* (NCS, 1993-1994).

6.3 PACKAGING AND SHIPPING

6.3.1 Distribution

The distribution effort for the 1994 NAEP Assessment involved packaging and mailing documents and associated forms and materials to the Westat supervisors for the main and trend assessments. The NAEP materials distribution system, initially developed by NCS in 1990 to control shipments to the schools and supervisors, was enhanced and utilized. Files in this system contained the names and addresses for shipment of materials, scheduled assessment dates, and a listing of all materials available for use by a participant in a particular assessment. Changes to any of this information were made directly in the file either manually or via file updates provided by Westat.

The bar code technology introduced by NCS in the 1990 assessments was again used for document control in 1994. A unique 10-digit number, consisting of the three-digit booklet number or form type, a six-digit sequential number, and a check digit, was used to identify each document. Each form was assigned a range of ID numbers. Bar codes reflecting this ID number were applied to the front cover of each document by NCS bar code technology and an ink jet printer. Later, as bar codes were read during the scanning process, the document ID number was incorporated into each student record.

The booklets were then spiraled into bundles, according to the design provided by ETS. Bundles of 11 booklets were created in the pattern dictated by the bundle maps. The booklets were arranged in such a manner that each booklet appeared first in a bundle approximately the same number of times and the booklets were evenly distributed across the bundles. This assured that sample sizes of individual booklets would not be jeopardized if entire bundles were not used. Since all Administration Schedules for each scheduled session were preprinted with the booklet IDs designated for that session, only bundles of 11 booklets were created. Three bundles of booklets were preassigned to each session, giving each 33 booklets. This number most closely approximated the average projected session size of 30 students and allowed extra booklets for either additional students or to replace defective booklets.

Initially, 6,251 individual session shipments were sent for the 1994 NAEP assessments. Approximately 800 additional shipments of booklets and miscellaneous materials were sent. All outbound shipments were recorded in the NCS outbound mail management system. Each address label had a bar code containing the school number. The bar code was read into the system, which determined the routing of the shipment and the charges. Information was recorded in a file on the system which, at the end of the day, was transferred to the mainframe by a personal computer. A computer program could then access information to produce reports on shipments sent, regardless of the carrier used.

A toll-free telephone line was maintained for supervisors and administrators to request additional assessment materials. To process a shipment, a clerk asked the caller information such as primary sampling unit (PSU), school ID, assessment type, city, state, and ZIP code. This information was then entered into the online short shipment system and a particular school and mailing address would be displayed on the screen to verify with the caller. The system allowed NCS staff to change the shipping address for individual requests. The clerk would proceed to the next screen which displayed the materials to be selected. After the clerk entered the requested items, the due date and the method of shipment, the system produced a packing list and mailing labels.

6.4 PROCESSING

6.4.1 Overview

The following describes the stages of work involved in receiving and processing the documents used in the 1994 assessment. NCS staff created a set of predetermined rules and specifications that were to be followed by the processing departments within NCS. Project staff performed a variety of procedures on materials received from the assessment supervisors before releasing these materials into the NCS NAEP processing system. Control systems were used to monitor and control all NAEP materials returned from the field. The NAEP process control system contained the status of all sampled schools for all sessions and their scheduled assessment dates. As materials were returned, the process control system was updated to indicate receipt dates, to record counts of materials returned, and to document any problems discovered in the shipments. As documents were processed, the system was updated to reflect the processed counts. NCS report programs allowed ETS, Westat, and NCS staff to monitor the progress in the receipt control operations.

An "alerts" process was utilized to record, monitor, and categorize all discrepant or problematic situations. Throughout the processing cycle, alert situations were identified based upon the processing specifications. These situations were either flagged by computer programs or identified using clerical procedures. All situations that could not be directly resolved by the staff involved in the given process were documented. A form describing the problem was completed and the information was forwarded to project personnel for resolution.

NCS's work flow management system was used to track batches of student booklets through each processing step, allowing project staff to monitor the status of all work in progress. The WFM system was also used by NCS to analyze the current work load, by project, across all work stations. By routinely monitoring this data, NCS's management staff was able to assign priorities to various components of the work and monitor all phases of the data receipt and processing.

NCS used a team approach to facilitate the flow of materials through all data processing steps. The image processing team checked in the materials from the field, created the batches to be scanned, scanned the booklets, edited the information when the program found errors or inconsistencies, selected quality control samples, and sent the completed batches to the warehouse for storage. The team approach provided the advantages of less duplication of effort and improved quality control measures.

6.4.2 Document Receipt

Shipments were to be returned to NCS packaged in the original boxes. As mentioned earlier, NCS packaging staff applied a bar code label to each box that indicated the NAEP school ID number. Magenta labels were used to identify long-term trend assessment shipments, while blue labels were used to identify main assessment shipments. When the shipment arrived at the NCS dock area, this bar code was scanned to a personal computer file and sorted by assessment type. The shipment was then forwarded to the receiving area. The file was uploaded to the mainframe and the shipment receipt date was applied to the appropriate school within the process control system. This provided the current status of receipts regardless of any processing delays. The receipt was reflected on the system status report provided to the receiving department and also supplied to Westat via electronic data file transfer.

The process control system could be updated manually to reflect changes. Receiving personnel also checked the shipment to verify that the contents of the box matched the school indicated on the label. Each shipment was checked for completeness and accuracy. Westat was notifiied of any shipment not received within seven days of the scheduled assessment date, and in some cases a trace was initiated.

The contents of the package were removed and separated by session, if multiple sessions were returned in one box. The packing list was used by the assessment supervisor to indicate all materials being returned to NCS. The shipment was checked to verify that all booklets preprinted or handwritten on the Administration Schedule were returned with the shipment and that all Administration Codes matched from the booklet covers to the Administration Schedule. If discrepancies were discovered at any step in this process, the receiving staff issued an alert and held the session for resolution by the NAEP project staff.

If a make-up session had been scheduled, receiving staff issued an information alert to facilitate tracking, and the documents were placed on holding shelves until the make-up session documents arrived.

6.4.3 Batching and Scanning Documents

Once all booklets listed on the Administration Schedule for sessions containing scannable documents were verified as being present, the entire session, both the Administration Schedule and booklets, was forwarded to the batching area and a batch was created on the work flow management system using the scannable Administration Schedule as a session header. The booklets were batched by grade level and assessment type. Each batch was assigned a unique number. The number, created on the image capture environment system and automatically uploaded to the work flow management system, facilitated the internal tracking of the batches and allowed departmental resource planning. All other scannable documents, questionnaires, and rosters were batched by document type in the same manner.

The batch was then forwarded to scanning where all information on the Administration Schedule and booklets was put through an image scanner. All information from the Administration Schedules was read by the intelligent character recognition engine and verified by online editing staff for correctness. Information gathered throughout this process, which included the school number, session code, counts of the students in the original sample, the supplemental sample, and the total sample; numbers of students withdrawn, excluded, to be assessed, absent, originally assessed, and assessed in makeup; and the total number of assessed students, was then transferred electronically to Westat weekly to produce participation statistics.

Approximately five percent of the booklets in 1994 were nonscannable. For all sessions containing nonscannable booklets—long-term trend reading/writing booklets—all booklets listed on the Administration Schedule were verified as being present. In addition, the administration codes were matched from the Administration Schedule to the booklet cover. The Administration Schedule was then grouped with schedules from other long-term trend reading/writing sessions and a batch was created on the work flow management system for these control documents. The Administration Schedule batch was then forwarded to scanning where all information on the Administration Schedule was scanned into the system and verified for correctness via online editing. The same information gathered through this process—school number, session code, counts of the students in original sample, etc.—was also transferred to Westat on a weekly basis to produce participation statistics. Nonscannable booklets from these sessions were then forwarded to the batching area.

6.4.4 Questionnaires

Three rosters were used to account for all questionnaires. The Roster of Questionnaires recorded the distribution and return of the school characteristics and policies questionnaires, the excluded student questionnaires, and the IEP/LEP student questionnaires. The Roster of Teacher Questionnaires - Grade 4, and the Roster of Teacher Questionnaires - Grade 8, recorded teacher questionnaires distributed and returned for their respective grades. Some questionnaires may not have been available for return with the shipment. These were returned to NCS at a later date in an envelope provided for that purpose. The questionnaires were submitted for scanning as sufficient quantities became available for batching.

Receipt of the questionnaires was entered into the system using the same process as was used for the Administration Schedules. The rosters were grouped with other rosters of the same type from other sessions, and a batch was created. The batch was then forwarded to scanning where all information on the rosters was scanned into the system.

6.4.5 Booklet Accountability

NCS used a sophisticated booklet accountability system to track all booklets distributed. Prior to the distribution of materials, unique booklet numbers were read into a file, by bundle. This file was then used to control distribution by assigning specific bundles to supervisors or schools. This assignment was recorded in the materials distribution system.

When shipments were received, the used booklets were submitted to processing. The unused booklets were batched and the booklet ID bar code was read into a file by the bar code scanner. This file and the processed documents file were later compared to the original bundle security file. A list of unmatched booklet IDs was printed in a report that was used to confirm nonreceipt of individual booklets. At the end of the assessment period, the supervisors returned all unused materials. When these materials were returned, the booklet IDs were read into a file by the bar code scanner. Any major discrepancies were directed to Westat for follow-up. The unused materials were then inventoried and sent to the NCS warehouse for storage.

6.4.6 Data Entry

The data entry process was the first point at which booklet level data were directly available to the computer system. Depending on the type of document, one of two methods was used to transcribe NAEP data to a computerized form. The data on scannable documents were collected using NCS optical scanning equipment and image-capture technology. Nonscannable materials were keyed through an interactive online system. In both cases, the data were edited and suspect cases were resolved before further processing.

The majority of student booklets, questionnaires, and control documents were scannable. Throughout all phases of processing, the student booklets were batched by grade and session type (reading or geography/U.S. history). The scannable documents were then transported to a slitting area where the folded and stapled spine was removed from the document. In addition to capturing the student responses, the bar code identification numbers used to maintain process control were decoded and transcribed to the NAEP computerized data file.

During the scanning process each scannable NAEP document was given a unique print-after-scan number consisting of the scan batch number and the sequential number within the batch. The number was assigned to and printed on one side of each sheet of each document as it exited the scanner. This permitted the data editors to quickly and accurately locate specific documents during the editing phase. The number remained with the data record and provided a method for easy identification and quick retrieval of any document.

The data values were captured from the booklet covers and Administration Schedules and were coded as numeric data. Unmarked fields were coded as blanks and processing staff were alerted to missing or uncoded critical data. Fields that had multiple marks were coded as asterisks (*). The data values for the item responses and scores were returned as numeric codes. The multiple-choice, single response format items were assigned codes depending on the position of the response alternative; that is, the first choice was assigned the code "1," the second "2," and so forth. The mark-all-that-apply items were given as many data fields as response alternatives; the marked choices were coded as "1" and the unmarked choices as blanks. The images of constructed response items were saved as a digitized computer file. The fields from unreadable pages were coded "X" as a flag for resolution staff to correct.

As the scanning program completed scanning each stack, the stack was removed from the output hopper and placed in the same order on the output cart. The next stack was removed from the cart, placed into the input hopper and the scanning resumed. When the operator had completed processing the last stack of the batch, the program was terminated. This closed the dataset and it automatically became available for the edit process. The scanned documents were then forwarded to a holding area in case they needed to be retrieved for resolution of edit errors.

NCS used the intelligent character recognition engine to read various hand and machine print on the front cover of the assessment and supervisor documents. Some information from scannable long-term trend booklets, the Administration Schedule, the Rosters of Questionnaires, and some questions in the school characteristics and policies questionnaire were read by the engine and verified by a key entry operator. Analysis by NCS development staff of the accuracy of characters read via intelligent character recognition determined that the engine read as well as two people processing information using a key entry and 100 percent verify method of data input. In all, the engine read 14.4 million characters—10.7 million handwritten characters and 3.7 million machine-printed characters. This saved NAEP field staff and school personnel a significant amount of time, since they no longer had to enter these data by gridding rows and columns of data.

To provide yet another quality check on the image scanning and scoring system, NCS staff implemented a quality-check process by creating labels with a valid score designated on them. Each unique item scored via the image system was quality check twice. Labels were attached to blank, unused booklets by clerical staff and sent through the scanning process. An example of the label used is given below.

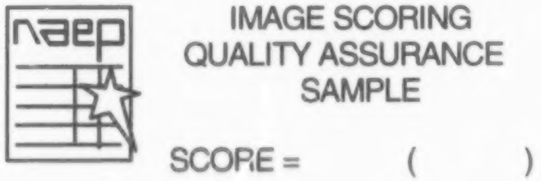

naep

IMAGE SCORING
QUALITY ASSURANCE
SAMPLE

SCORE = ( )

Although the quality-control booklets were batched and processed separately from assessed student booklets, they were sent through the same process as the student documents. Since all of a specific item are batched together for transmission to the scoring facility, the quality-control responses were integrated with and transmitted simultaneously to the scoring facility with the student responses. During the scoring process, both student responses and the quality-control items were randomly displayed so scores could be applied.

When a reader saw the quality-control label on the monitor, he or she notified the team leader to watch and confirm the scoring while the reader assigned the score on the label. The quality-control booklets were included in the pool of all items to be drawn from for the 25 percent reliability rescore. Analysis of the data captured from this quality-assurance process is addressed more fully in Chapter 8.

A process of key entry and verification was used to make corrections to the nonscannable long-term trend reading/writing booklets and the teacher, IEP/LEP, and excluded student questionnaires. The Falcon system that was used to enter this data is an online data entry system designed to replace most methods of data input such as keypunch, key-to-disk, and many of the microcomputer data entry systems. The terminal screens were designed for operator speed and convenience. The fields to be entered were titled to reflect the actual source document. Thus, all key entry fields were specific to the NAEP student documents or questionnaire types being keyed.

To do the proper edits, a detailed document definition procedure was designed which allowed NCS to define an item once and use it in many blocks and to define a block once and use it in many documents. The procedure used was a *document* file that pointed to the appropriate blocks on a *block* file which in turn pointed to appropriate items on an *item* file. With this method of definition, a document was made up of blocks, which were made up of items.

6.4.7 Data Validation

Each dataset produced by the scanning system contained data for a particular batch. These data had to be edited for type and range of response. The data entry and resolution system used was able to process a variety of materials from all age groups, subject areas, control documents, and questionnaires simultaneously, as the materials were submitted to the system from scannable and nonscannable media.

The data records in the scan file were organized in the same order in which the paper materials were processed by the scanner. A record for each batch header preceded all data records for that batch. The document code field on each record distinguished the header record from the data records.

When a batch header record was read, a pre-edit data file and an edit log was generated. As the program processed each record within a batch from the scan file, it wrote the edited and reformatted data records to the pre-edit data file and recorded all errors on the edit log. The data fields on an edit log record identified each data problem by the batch sequence number, booklet serial number, section or block code, field name or item number, and data value. After each batch had been processed, the program generated a listing or online edit file of the data problems and resolution guidelines. An edit-log listing was printed at the termination of the program for all non-image documents. Image "clips" were then routed to online editing stations for those documents that were image-scanned.

As the program processed each data record, it first read the booklet number and checked it against the session code for appropriate session type. Any mismatch was recorded on the error log and processing continued. The booklet number was then compared against the first three digits of the student identification number. If they did not match, a message was written on the error log. The remaining booklet cover fields were read and validated for the correct range of values. The school codes had to be identical to those on the process control system record. All data values that were out of range were read "as is" but flagged as suspect. All data fields that were read as asterisks were recorded on the edit log or online edit file.

Document definition files described each document as a series of blocks, which in turn were described as a series of items. The blocks in a document were transcribed in the order that they appeared in the document. Each block's fields were validated during this process. If a document contained suspect fields, the cover information was recorded on the edit log, along with a description of the suspect data. The edited booklet cover was transferred to an output buffer area within the program. As the program processed each block of data from the dataset record, it appended the edited data fields to the data already in this buffer.

The program then cycled through the data area corresponding to the item blocks. The task of translating, validating, and reporting errors for each data field in each block was performed by a routine that required only the block identification code and the string of input data. This routine had access to a block definition file that had, for each block, the number of fields to be processed, and, for each field, the field type (alphabetic or numeric), the field width in the data record, and the valid range of values. The routine then processed each field in sequence order, performing the necessary translation, validation, and reporting tasks.

When the entire document had been processed, the completed string of data was written to the data file. When the program encountered the end of a file, it closed the dataset and generated an edit listing for non-image and key-entered documents. Image scanned items that required correction were displayed on an online editing terminal.

Accuracy checks were performed on each non-image batch. The record of every 500th document of each booklet/document type was printed in its entirety, with a minimum of one document type per batch.

This record was checked, item by item, against the source document for errors. If errors were found, project personnel were contacted and processing stopped.

6.4.8 Editing Non-Image and Key-Entered Documents

Throughout the system, quality procedures and software ensured that the NAEP data were correct. The machine edits performed during data capture verified that each sheet of each document was present and that each field had an appropriate value. All batches entered into the system, whether key entered or machine scanned, were checked for completeness.

Data editing took place after these checks. This consisted of a computerized edit review of each respondent's record and the clerical edits necessary to make corrections based upon the computer edit. This data editing step was repeated until all data fell within a valid range.

The first phase of data editing was designed to validate the population and ensure that all documents were present. A computerized edit list, produced after NAEP documents were scanned or key entered, and all the supporting documentation sent from the field were used to perform the edit function. The hard copy edit list contained all the vital statistics about the batch. The number of students, school code, type of document, assessment code, error rates, suspect cases, and record serial numbers were among these elements. Using these inputs, the data editor verified that the batch had been assembled correctly and each school number was correct.

During data entry, counts of processed documents were generated by type. These counts were balanced against the information captured from the administration schedules. The number of assessed and absent students processed had to match the numbers indicated on the process control system.

In the second phase of data editing, an experienced editing staff used a predetermined set of specifications to review the field errors and record any necessary correction to the student data file. The same computerized edit list used in the first phase was used to perform this function. The process was as follows:

The editing staff reviewed the edit log prepared by the computer and the area of the source document that was noted as being "suspect" or containing possible errors. The current composition of the field was shown in the edit box. The editing staff checked this piece of information against the NAEP source document. At that point, one of the following took place:

> *Correctable error*: If the error was correctable by the editing staff according to the editing specifications, the corrections were noted on the edit log.
>
> *Alert*: If an error was not correctable according to the specifications, an alert was issued to the operations coordinator for resolution. Once the correct information was obtained, the correction was noted on the edit log.
>
> *Noncorrectable error*: If a suspected error was found to be correct as stated and no alteration was possible according to the source document and specifications, the programs were tailored to allow this information to be accepted into the data record and no corrective action was taken.

The corrected edit log was then forwarded to the key entry staff for processing. When all corrections were entered and verified for a batch, an extract program pulled the corrected records into a mainframe dataset. At this point, the mainframe edit program was initiated. The edit criteria were again applied to all records. If there were further errors, a new edit listing was printed and the cycle began again.

When the edit process had produced an error-free file, the booklet ID number was posted to the NAEP tracking file by age, assessment, and school. This permitted NCS staff to monitor the NAEP processing effort by accurately measuring the number of documents processed by form. The posting of booklet IDs also ensured that a booklet ID was not processed more than once.

6.4.9 Data Validation and Editing of Image-Processed Documents

The paper edit log was replaced by online viewing of suspect data for all image-processed documents. The edit criteria for each item or items in question also appeared on the screen at the same time the suspect item was displayed for rapid resolution. Corrections were made at this time. The system employed an edit/verify system which ultimately enabled two different online-edit operators to view the same suspect data and work on it separately. The "verifier" must make sure that the two responses (one from either the "entry" operator or the intelligent character recognition engine) were the same before the system would accept that item as being corrected. The verifier was able to overrule or agree with the original correction made if the two were discrepant. If the editor was unable to determine the appropriate response, he or she escalated the suspect situation to a supervisor.

When an entire batch was through the edit phase, it was then eligible for the count verification phase. The administration schedule data were examined systematically for booklet IDs that should have been processed (assessed, absent, and excluded administration codes). The documents under an individual administration schedule were then inspected to ensure that all of the booklet IDs listed on it were present.

With the satisfactory conclusion of the count verification phase, the edited batch file was uploaded to the mainframe where it went through yet another edit process. A paper edit log was then produced, and, if errors remained, the paper edit log was forwarded to another editor. When this edit was satisfied, the appropriate tracking mechanisms (the process control and work flow management systems) were updated.

6.4.10 Data Transmission

Due to the rapid pace of scoring on an item-by-item basis, the NCS scoring specialists found it necessary to continually monitor the status of work available to the readers and plan the scoring schedule several weeks in advance. On Wednesday of each week, the NCS scoring specialist planned the schedule for the next two weeks. That information was then provided to the person in charge of downloading data to the scoring center. By planning the scoring schedule two weeks in advance, the scoring specialists were able to ensure that readers would have sufficient work for at least one week, after which the next download would occur to supplement the volume of any unscored items and add an additional week's work to the pool of items to score. Additionally, by scheduling two weeks' data for transmission, flexibility was added to the scoring schedule, making it possible to implement last minute changes in the schedule once the items had been delivered to the scoring center. Depending on the number of items to be transmitted, the actual scheduling was conducted on Friday or divided into two smaller sessions on Thursday and Friday.

Delivery of data to the scoring center—located approximately five miles from NCS's main facility in Iowa City—was accomplished via several T1 transmission lines linking the mainframe computers and the NAEP servers at the site of document scanning in the main facility, with the scoring servers dedicated to distributing work to the professional readers at the scoring center. The actual task of scheduling items for downloading was accomplished using code written by the image software development team. This code enabled the person scheduling the download to choose a team of readers and select the scheduled items from a list of all items that team would be scoring throughout the scoring project. This process was repeated for all teams of readers until all anticipated work was scheduled. Once this task was completed, the scheduled job was tested to determine if sufficient free disk space existed on the servers at the scoring center. If, for any reason, sufficient disk space was not available, scheduled items could be deleted from the batch individually or as a group until the scheduled batch job could accommodate all items on the available disk space at the scoring center. Once it was determined that there was sufficient disk space, transmission of student responses commenced. Data transmission was typically accomplished during off-shift hours to minimize the impact on the system's load capacity.

6.5 DATA DELIVERY

The 1994 NAEP data collection resulted in several classes of data files—student, school, teacher, excluded student, IEP/LEP student, sampling weight, student/teacher match and item information. Item information included item data from all assessed students in 1994, item data for the short-term reading trend, and item data from the special study comparing image-based and paper-based scoring. Data resolution activities occurred prior to the submission of data files to ETS and Westat to resolve any irregularities that existed. This section details additional steps performed before creating of the final data files to ensure the most complete and accurate information was captured.

An important quality control component of the image scoring system was the inclusion, for purposes of file identification, of an exact copy of the entire student edit record, including the student booklet ID number, with every image of a student's response to a constructed-response item. These edit files also remained in the main data files residing on the NCS mainframe computer. By doing this, exact matching of scores assigned to constructed-response items and the rest of each individual student's data was guaranteed as the booklet ID for each image was part of every image file.

When all the responses for an individual item had been scored, the system automatically submitted all item scores assigned during scoring and their edit records to a queue to be transmitted to the mainframe. Project staff then initiated a system job to transmit all scoring data to be matched with the original student records on the mainframe. A custom edit program matched the edit records of the scoring files to those of the original edit records on the mainframe. As matches were confirmed, the scores were applied to those individual files. After completion of this stage, all data collected for an individual student was located in one single and complete record/file identified by the edit record.

Some of the assessed students were determined to be ineligible for the assessment because they did not match the particular age/grade being sampled or because of unusual circumstances. At the conclusion of each assessment, it was necessary to delete the records of these students from the NAEP database. Deleting this information required compiling a list of all student records that had been processed with administration codes other than those for assessed students. To do this, the process control system and the Administration Schedule data were referenced. If the system showed a discrepancy, project personnel pulled the Administration Schedules and other documentation (e.g., alerts, student booklets, etc.) to verify and resolve the discrepancy.

The edits and data verification performed on the IEP/LEP student questionnaires assured that information regarding the IEP/LEP status of the students was not left blank. If there was no indication as to IEP or LEP on the questionnaire cover, the edit clerk cross-checked the administration schedule(s) and student booklet cover to confirm the IEP/LEP status of the student. If this information was not available from the questionnaire cover, booklet cover, or the administration schedule, the edit clerk viewed the information indicated in question #1 (which asked why the student was classified IEP or LEP) to see whether responses written there might yield useful information. Then the determination was made as to how the student should be classified.

The excluded student questionnaires were used for the long-term trend assessments. As with the IEP/LEP questionnaires, if data regarding the reason for exclusion was missing or inconsistent among the various documents, the edit clerk crosschecked the documents to ensure that the most complete and accurate was captured.

The school questionnaires were revised for 1994 so that some items that had required school staff to provide a percentage figure by gridding ovals in a matrix were changed to allow the respondent to simply write the percentage in a box. These data were then captured via intelligent character recognition technology and verified by an edit operator.

To obtain the best possible match of teacher questionnaires to student records, the same processes that were followed in 1992 were refined in 1994. The first step in matching was to identify teacher questionnaires that had not been returned to NCS for processing, so as not to include the students of these teachers from the matching process. Student identification numbers that were not matched to a teacher questionnaire were then crossreferenced with the corresponding Administration Schedule and Roster of Teacher Questionnaires to verify the teacher number, teacher period, and questionnaire number recorded on these control documents. If a change could be made that would result in a match, the correction was applied to the student record. The NAEP school numbers listed on the Roster of Questionnaires, Administration Schedule, and teacher questionnaire were verified and corrected, if necessary.

Once these resolutions were made, any duplicate teacher numbers that existed within a school were crossreferenced with the Rosters of Questionnaires for resolution, if possible. In one jurisdiction that had multiple sessions in many schools, a number of the schools used a single Roster of Questionnaires for each session. This resulted in a larger than expected number of duplicate teacher numbers that could not be resolved. The overall quality of the matching process improved in 1994 as a result of the inclusion of the teacher number and period on the Administration Schedule. Since this information was located together on a single, central control document, the ability to match and resolve discrepant or missing fields was simplified.

After all data processing activities were completed, data cartridges or tapes were created and shipped via overnight delivery to ETS and/or Westat, as appropriate. A duplicate archive file is maintained at NCS for security/backup purposes.

6.6 MISCELLANEOUS

6.6.1 Storage of Documents

After the batches of image-scanned documents had successfully passed the editing process, they were sent to the warehouse for storage. Batches of 1992 rescore booklets were sent to the scoring area after passing the edit phase of processing, because they were also to be scored on paper. Once paper scoring was completed, 1992 rescore booklets were also sent to the warehouse for storage. The storage locations of all documents were recorded on the inventory control system. Unused materials were sent to temporary storage to await completion of the entire assessment. After the data tape was accepted, extra inventory was destroyed and a nominal supply of materials was stored permanently.

6.6.2 Quality Control Documents

ETS requested that a random sample of booklets and the corresponding scores/scoring sheets be pulled for an additional quality control check. Because no scoring sheet was available for image-scanned documents, ETS used scores sent to them on a data tape to verify the accuracy of applied scores. For nonscannable trend booklets and for the 1992 rescore booklets that were scored on paper, both the booklet and its corresponding score sheet were sent to ETS. An average of 20 of each booklet and scores/scoring sheets for each document type were selected at random by NCS. All of these documents were selected prior to sending the booklets to storage and were then sent to ETS to verify the accuracy and completeness of the data.

6.6.3 Alert Analysis

Even though Receiving Department personnel were trained in the resolution of many problematic situations, some problems required resolution by NAEP staff. The types of problems were categorized and codes were assigned. For any unusual situations, Westat was called so that the Assessment Supervisors could be notified immediately to avoid further problems in test administration.

Many discrepancies were found in the receiving process that did not require an alert to be issued, but did require a great deal of effort to resolve in order to provide the most complete and accurate information. These included blank fields on covers of booklets as well as discrepancies between the booklet covers and the administration schedule. There were a total of 281 alerts for the national assessment.

Chapter 7

PROFESSIONAL SCORING[1]

Claudia A. Gentile
Educational Testing Service

Charles Brungardt and Patrick B. Bourgeacq
National Computer Systems

7.1 INTRODUCTION

Scoring of the 1994 NAEP constructed-response items was conducted primarily on the NCS Computerized Image scanning and scoring system. With the exception of a few items that did not lend themselves to image scanning (colored-pencil items, for example) and those responses from previous years' administrations that were rescored as part of trend studies, all responses were scored by readers working online at image stations.

One of the main advantages of image technology for NAEP scoring was in the area of sorting and distributing work to scorers. All student responses for a particular item, regardless of where spiraling had placed that item in the various booklet forms, were batched together for presentation to a team of readers. This allowed trainers to conduct training on one item at a time, rather than a block of related items, effectively focusing readers' attention on the complexities of a single item.

Quality Control Measures. A number of tools built into the system allowed table leaders and trainers to continuously monitor reader performance. The system automatically routed 25 percent of student responses to other members of the team for second scoring. Readers were given no indication of whether the response had been scored by another reader, thereby making second scoring truly blind. On-demand, real-time reports on interreader reliability (drawn from those items that were second-scored) presented extremely valuable information on team and individual scoring. Information on adjacent and perfect agreement, score distribution, and quantity of responses scored were continuously available for consultation.

Similarly, back-reading of student responses could be accomplished in an efficient and timely manner. Table leaders were able to read a large percentage of responses, evaluating the appropriateness and accuracy of the scores assigned by readers on their teams.

In addition, project management tools assisted table leaders in making well-informed decisions. For example, knowledge of the precise number of responses remaining to be scored for a particular item allowed table leaders to determine the least disruptive times for lunch breaks.

---

[1] Claudia A. Gentile is the director of scoring for all the NAEP curriculum areas. Patrick B. Bourgeacq is the NCS/NAEP scoring project director. Charles Brungardt coordinates communications between NCS's performance scoring center and the operations departments.

Measures were taken to reduce possible reader fatigue or other problems that might result from working continuously at a computer terminal. Readers took hourly breaks, in addition to their lunch break, to reduce the degree of visual fatigue.

Scope of Work. Readers were grouped in teams of 6 to 10 readers per team. Individual rooms were set up with each room containing teams for a single subject area. Tables 7-1, 7-2, and 7-3 contain the counts of individual types of constructed-response items scored and the number of readers and table leaders for the 1994 assessments.

Table 7-1
Number of Constructed-Response Items
Long-Term Trend Assessments

Subject	Age 9	Age 13	Age 17	Total
Reading	5	8	9	22
Writing	6	6	6	18
Mathematics	44	27	57	128
Total	55	41	72	168

Table 7-2
Number of Constructed-Response Items
1993-94 Main Assessments

U.S. HISTORY				
Grade	**Short Constructed-Response 2-point***	**Short Constructed-Response 3-point**	**Extended Constructed-Response 4-point**	**Total**
4	0	16	5	21
4/8	1	8	2	11
8	0	21	6	27
8/12	0	7	2	9
12	1	26	16	43
Total	2	78	31	111

**Note*: These items were dichotomously scored.*

Table 7-2, cont.

Number of Constructed-Response Items
1993-94 Main Assessments

GEOGRAPHY				
Grade	**Short Constructed-Response 2-point***	**Short Constructed-Response 3-point**	**Extended Constructed-Response 4-point**	**Total**
4	2	13	6	21
4/8	1	7	2	10
8	1	16	3	20
8/12	0	8	3	11
12	0	15	12	27
Total	4	59	26	89

**Note*: These items were dichotomously scored.*

READING				
Grade	**Short Constructed-Response 2-point***	**Short Constructed-Response 3-point**	**Extended Constructed-Response 4-point**	**Total**
4	20	9	6	35
4/8	8	0	2	10
8	11	15	6	32
8/12	16	6	5	27
12	22	19	8	49
NAEP Reader Grade 8	8	0	3	11
NAEP Reader Grade 12	8	0	3	11
Total	93	49	33	175

**Note*: These items were dichotomously scored.*

BEST COPY AVAILABLE

Table 7-3
1993-94 NAEP Assessment
Professional Scoring - Readers and Dates

Assessment	Number of Table Leaders	Number of Readers	Dates
Long-Term Trend			
Age 9 (Winter)	1	4	02/07/94 - 04/01/94
Age 13 (Fall)	1	5	11/02/93 - 01/18/94
Age 17 (Spring)	1	4	04/11/94 - 06/10/94
Writing Holistic	5	41	07/11/94 - 07/15/94
Writing Mechanics	4	21	07/18/94 - 08/18/94
Geography	2	15	03/28/94 - 05/23/94
U.S. History	3	19	03/28/94 - 05/24/94
Reading			
Main	16	147	03/28/94 - 06/10/94
NAEP Reader	1	13	03/28/94 - 06/16/94

7.2 LONG-TERM TREND ASSESSMENTS

7.2.1 Characteristics of the Scoring Guides

Writing. All of the writing items for the three long-term trend assessments (fall–age 9, winter–age 13, and spring–age 17) were scored using the primary trait method. This method focused on the writer's effectiveness in accomplishing specific assigned tasks. The primary trait scoring criteria defined five levels of task accomplishment: not rated, unsatisfactory, minimal, adequate, and elaborated. The scoring standard for each item described these levels in detail. Some of these items were also scored for secondary traits, which involved indicating the presence or absence of elements that were of special significance to a particular item (i.e., whether notes were made before writing or whether critical information was filled out on a form). The scoring guides for the constructed-response writing items focused on students' abilities to write in informative, persuasive, and narrative manners. The guides for the writing items were based on a range of scores denoting unsatisfactory writing to address the task, minimal writing to address the task, satisfactory writing to address the task, and elaborated writing to address the task.

Reading. The scoring guides for the constructed-response reading items focused on students' abilities to perform various reading tasks: identifying the author's message or mood and substantiating their interpretation, making predictions based on given details, and comparing and contrasting. The guides for the reading items varied somewhat, but typically included a range of scores denoting inability to address the task, unsatisfactory responses, minimal ability in accomplishing the task, satisfactory ability in addressing the task, or elaborated responses addressing the task fully. Some of the reading items received secondary scoring based on what reactions or information the student gave (i.e., whether the response was mostly content based, form based, a subjective reaction, or some combination of the three).

Mathematics. The trend mathematics items were scored on a right/wrong basis. The scoring criteria identified the correct or acceptable answers for each item in each block. The scores for these items included a 0 for no response, a 1 for a correct answer, or a 2 for an incorrect or "I don't know" response. Certain items in the trend mathematics booklets were scored only for attemptedness. This scoring consisted of merely checking to see whether the student had responded in any way to that item, in which case the item was determined to have been reached or attempted. The scoring here was 0 for not having reached the item (blank) or 1 for any writing in the space provided.

Science. The science assessment had no constructed-response items.

7.2.2 Training

Training preparation for the three long-term trend scoring projects included covering over the scores assigned in previously scored assessment booklets from 1984 (reading responses) and 1988 (writing responses) to ensure that scoring for training, and subsequent trend reliability scoring, would be done without knowledge of the previous scores given.

The formal training for the trend assessments was divided into two parts to accommodate the reading and writing items. The trend scoring project was comprised of one team of five readers with one team leader. During training each reader received a photocopied packet of materials used in the 1984 scoring of the reading items and the 1988 scoring of the writing items.

Because the mathematics items were scored as right, wrong, or omitted, lengthy training for scoring these items was unnecessary. In an orientation period, clerical staff were trained to follow the procedures for scoring the mathematics items and became familiar with the scoring standards, which listed the correct answer for the items in each of the blocks.

Prior to scoring any 1994 reading and writing trend material, a training reliability report was generated using a 25 percent sampling of the 1984 and 1988 assessment materials. Following the formal training sessions, the readers scored this material on scannable scoring sheets produced for specific booklet types with the appropriate trend items preprinted on the scoring sheets. These sheets were then routed to scanning under a special job number to ensure that this material was labeled for training scoring only. The project coordinator was able to generate a computer report that listed the individual and group percent agreement by item. The system automatically compared the new score with the score assigned in the 1984 or 1988 process and produced a report on the training reliability. The NCS scoring specialist then conferred with the appropriate ETS staff on this training reliability agreement report before proceeding with scoring.

7.2.3 Scoring

Reliability studies were conducted for the scoring of the trend reading and writing items. For the 1994 material, 33 percent of the constructed-response items were scored by a second reader to produce interreader reliability statistics. In addition, a trend reliability study was conducted to ensure that the scoring procedures were consistent with those used in 1984 and 1988. Twenty percent of the 1984 reading responses and 25 percent of the 1988 writing responses were sampled. The scoring of these trend samples was intermixed with the scoring of the 1994 reading and writing trend material. The readers selected a bundle of approximately 25 booklets and gridded their scores on separate scannable scoring sheets for each item. These sheets were then scanned and cross-referenced with the original data tape to extract information for trend reliability reporting. Table 7-4 lists the number of unique items for each long-term trend component. The appropriate measures of reliability for scores from the current assessment are provided in Appendix H.

Table 7-4
1993-94 NAEP Long-Term Trend Interreader Reliability Ranges

Assessment	Number of Unique Items	Number of Items in Percentage Agreement Range				
		50-59%	60-69%	70-79%	80-89%	90-100%
Long-Term Trend						
Age 9 (Winter)						
Reading - 1994	5	0	0	0	0	5
Writing - 1994	6	0	0	0	3	3
Reading - Rescore	5	0	0	0	1	4
Writing - Rescore	6	0	0	0	2	4
Age 13 (Fall)						
Reading - 1994	7	0	0	1	6	0
Writing - 1994	6	0	0	1	2	3
Reading - Rescore	7	0	0	0	7	0
Writing - Rescore	6	0	0	0	4	2
Age 17 (Spring)						
Reading - 1994	8	0	0	1	5	2
Writing - 1994	6	0	0	1	5	0
Reading - Rescore	8	0	0	0	5	3
Writing - Rescore	6	0	0	0	3	3

The trend mathematics items were scored by trained clerical staff as part of the online editing procedures. The student response to each of the constructed-response items in the mathematics booklets was 'clipped' and displayed on an image terminal for scoring. The valid score points for each particular item were displayed on the screen, and the appropriate score was keyed. Each item was scored independently by two scorers. In the rare instance that a response received two different scores, the image system alerted the scorers to the discrepancy by displaying the response along with the two discrepant scores on the screen. The discrepancy was then resolved by the table leader.

7.2.4 Holistic Writing Scoring

To gain more information about how student writing performance changes across time at each grade level, NAEP conducted a holistic scoring of a subsample of the writing prompts included in the trend assessment. Holistic scoring evaluates the overall fluency of students' writing.

In the 1994 NAEP holistic scoring session, two items were scored from each of the three grade levels of the long-term trend writing assessment. The responses scored at grade 4 were from the items Flashlight, an imaginative task, and Spaceship, a persuasive task. At grades 8 and 11, the two items scored were Food on the Frontier, an informative task, and Recreation Opportunities, a persuasive task. The responses scored were taken from five assessment years: 1984, 1988, 1990, 1992, and 1994. Forty-one readers participated in the 1994 holistic scoring session. They were organized into five teams – four teams of eight and one team of nine readers – with each team led by an experienced team leader and assisted by one clerical aide.

The holistic writing trend scoring included a second scoring of 25 percent of all responses from each assessment year (1994, 1992, 1990, 1988, and 1994). Reliability reports were generated daily for each item being scored. These reports listed exact score agreement and adjacent score agreement between first and second scorings by individual reader and in total. The range of scores was from 0 to 6.

The readers were trained and scored the items in the following order:

Informative:	Food on the Frontier	Grade 11 followed by Grade 8
Persuasive:	Recreation Opportunities ...	Grade 11 followed by Grade 8
	Spaceship	Grade 4
Narrative:	Flashlight	Grade 4

The NCS scoring system generated a report that identified the interreader agreement by item. This report shows the number of times each item was read and the number of times each had second readings. The percent of exact reader agreement was listed, as well as the percent agreement of scores within one score point of each other. The ranges of interreader reliability for each item can be found in Table 7-5.

BEST COPY AVAILABLE

Table 7-5
1993-94 NAEP Assessment
Professional Scoring - Holistic Writing Scoring Reliability Figures

Grade	Item	Exact Agreement	Adjacent Agreement
4	Flashlight	59.7%	96.7%
	Spaceship	68.3%	98.0%
8	Food on the Frontier	58.2%	94.9%
	Recreational Opportunities	58.9%	95.9%
11	Food on the Frontier	57.5%	95.0%
	Recreational Opportunities	56.5%	95.5%
	Total	59.9%	96.0%

7.2.5 Mechanics Scoring

A subsample of the papers scored holistically received a further mechanics analysis. Students' responses at grade 4 to the Flashlight task and at grades 8 and 11 to the Recreation Opportunities task received a four-stage mechanics scoring.

At stage one, sentences were classified as either simple, complex, or compound. Sentence-level errors (sentence fragments and run-on sentences) were also coded at this stage. Stage two involved parsing students' text for phrase-level errors such as agreement errors and awkward constructions. The third stage focused on word-level errors (misspellings and word choice errors). In stage four, errors in capitalization and punctuation were coded. These codes were written directly onto photocopies of students' responses.

Every student response was scored by two readers. All disagreements were resolved by a third reader. The final coded responses were then typed into the database by a clerical staff member for further analysis (including a count of the number of words, number of sentences, and length of sentences for each response, and the average).

7.3 MAIN NAEP ASSESSMENT - OVERVIEW

7.3.1 Preparation for Training

A pool of papers to be used for training for main NAEP was selected by NCS staff during February 1994.

The training set for each short (2- or 3-point) item included:

Anchor Papers	10
Practice Papers	20
Calibration Set #1	5
Calibration Set #2	5
TOTAL	40 papers

The training set for each extended (4-point) item included:

Anchor Papers	15
Practice Papers	40
2 Qualification Sets	10 each
Calibration Set #1	5
Calibration Set #2	5
TOTAL	85 papers

To ensure that the ETS test development specialist would have a wide range of student responses from which to develop the master training set, NCS personnel copied approximately 100 papers for each 2- or 3-point item and 200 papers for each 4-point item. To ensure that training papers represented the range of responses obtained from the sample population, NCS personnel selected papers randomly from across the sample.

7.3.2 General Training Guidelines

ETS personnel conducted training for the constructed-response items on an item-by-item basis, so that each item could be scored immediately after training. Training and scoring occurred sequentially for reading items that were tied to a common stimulus. U.S. history and geography items were scored independently of the other items in their block.

The readers were divided into teams by subject area. All three subject areas followed a basic ETS plan for training. (Differences specific to the individual subject areas are discussed in the sections for those subjects.) Each member of a team received a copy of the stimulus and training materials for the items that his/her team would be scoring. Before training, each team member read the stimulus and discussed it under the guidance of the trainer where applicable. Next, the trainer presented and explained the anchor papers exemplifying the various score point levels. The team proceeded with each member scoring the practice papers and then discussing those papers as a group while the trainer clarified issues and answered questions. The papers selected for each training set were chosen to illustrate a range from easily classifiable responses to borderline responses for each score point.

When the trainer was satisfied that the readers were ready to begin scoring short constructed responses, the table leader signaled the system to release the responses to the team members who had successfully completed training. For extended constructed-response items, each team member was given a qualifying set that had been prescored by the trainer in conjunction with the table leader. Readers were required to score an exact match on 80 percent of the items in order to qualify for scoring. If a reader failed on the first attempt, the trainer discussed the discrepant scores with the reader and administered a second qualifying set. Again, 80 percent exact agreement was required to score the item. During the beginning stages of scoring, the team members discussed student responses with the trainer and table leader to ensure that issues not addressed in training were handled in the same manner by all team members.

After the initial training was completed, readers scored the items, addressing questions to the table leader and/or trainer when appropriate. Depending upon n-counts, length of responses, and complexity of the rubric, scoring time for all responses to an individual item ranged anywhere from half an hour to two weeks. Whenever a break longer than 15 minutes occurred in scoring, each team member received a set of calibration papers that had been prescored by the trainer and table leader. Each team member scored the calibration set individually, and then the team discussed the papers to ensure that scorer drift did not occur.

7.3.3 Qualification and Reliability Reports

Among the many advantages of the image scoring system are the ease with which workflow to readers can be regulated and scoring can be monitored. One of the utilities at a table leader's disposal was a qualification algorithm that was executed upon completion of training on an extended constructed-response item. At that time, a table leader passed out a qualification packet of 10 papers whose scores had been entered as a master key on the table leader's workstation. Upon completion of the packet, the table leader entered each reader's scores into the computer for tabulation and the computer calculated each reader's percent of exact, adjacent, and nonadjacent agreement with the master key. If a reader had a percent of exact agreement above a predetermined threshold, the reader was awarded a 'diploma' and allowed to begin scoring. Readers not reaching the predetermined threshold were handled on a case-by-case basis—typically receiving individual training by the ETS trainer or the NCS table leader before being allowed to begin scoring. A table leader also had the capability to cancel a reader's qualification to score an item if review of a reader's work had indicated that the reader was scoring inaccurately.

After scoring had commenced, review of each reader's progress was conducted using a back-reading utility that allowed a table leader to review every paper scored by each reader on the table. Typically a table leader would choose the ID number of a reader and review a minimum of 10 percent of the responses scored by that reader, making certain to note the score the reader awarded each response as well as the score a second reader gave that same paper as an interreader reliability check. Alternately, a table leader could select to review all responses receiving a specific score in order to determine if the whole team was scoring consistently. Both review methods utilized the same display screen and revealed the ID number of the reader and the score awarded. If the table leader disagreed with the score given a response, the table leader would discuss the discrepancy with the reader and possibly replace the score of the questionable response. The replacement of scores by the table leader was done only when the scorer had made an obvious error. The main purpose of this monitoring was to provide early identification of problems and opportunities to retrain scorers when needed.

During the scoring of an item, the table leader had the ability to monitor progress using an interreader reliability tool. This display tool could be used in either of two modes—to display information of all first readings versus all second readings, or to display all first reading of an individual versus all second readings of that individual. The information was displayed as a matrix with scores awarded during first readings displayed in rows and scores awarded during second readings displayed in columns. In this format, instances of exact agreement fell along the diagonal of the matrix. For completeness, data in each cell of the matrix contained the number and percentage of cases of agreement (or disagreement). The display also contained information on the total number of second readings and the overall percentage of reliability on the item.

The table leaders were able to monitor workflow using a status tool that displayed the number of items that had been completed, the number of items that had been first scored and still needed second scoring, and the number of items that had not been scored up to that time.

This chapter provides information about the reliability statistics available at the time of scoring. Further information about the reliability of main assessment items, including percent exact agreement, Cohen's (1968) Kappa, and intraclass correlations are provided in the subject analysis chapters (Chapters 12, 13, and 14) and in Appendix H.

7.3.4 Main U.S. History Assessment

7.3.4.1 Characteristics of the Scoring Guides

The U.S. history portion of the 1994 main assessment included a total of 111 discrete constructed-response items. The assessment used a variety of constructed-response items in order to measure different elements of students' understanding of American history. Each constructed-response item had a unique scoring guide that identified the range of possible scores for the item and defined the criteria to be used in evaluating student response. Two (time line) items were scored dichotomously. Seventy-eight items were scored on a 3-point scale with the option of giving the response a partially correct score. The 31 remaining items were extended constructed-responses and were scored on a 4-point scale.

2-Point (Dichotomous) Guide:
1 = Incorrect
3 = Correct

3-Point Guide:
1 = Inappropriate
2 = Partial
3 = Appropriate

4-Point Guide:
1 = Inappropriate
2 = Partial
3 = Essential
4 = Complete

The 19 U.S. history readers and three table leaders were organized into three teams, roughly by grade level. Each team was trained on a specific item and then scored that item in its entirety, usually during a single day. During the scoring, via the NCS image system, the table leader and the trainer monitored the individual readers via backreading and the various reliability reports.

7.3.4.2 Training

Training and scoring were done on an item-by-item basis. An item was scored using the ETS procedures described earlier.

7.3.4.3 Scoring

The formal scoring process on the image system began after the trainer was satisfied that the readers understood the scoring guide and could consistently apply it to the individual student response. Scoring usually began with the team gathering around a single terminal and jointly scoring about 25 responses. This helped facilitate discussion among the group members. Readers then returned to their own stations to begin scoring individually.

The table leader monitored the scoring of the item through back-reading, team and individual reliability reports (available online), and discussions with the reader and the team. Only one team worked on a particular item. If the table leader saw that the team had drifted from the scoring guide, scoring was stopped and discussion and group scoring took place until the table leader again felt comfortable with the direction of the scoring. The ETS trainer was available for more extensive training if needed.

A minimum of 25 percent of the U.S. history items were scored a second time. The image system presented all responses in the same manner, so the reader could not discern which responses were being scored for the first time and which were being scored for the second time. The table leader and the ETS trainer were able to monitor these figures at any point in the scoring. The system showed the overall reliability for the group scoring the item and individual reliability of the qualified readers. These reliability rates are shown in the table below.

Table 7-6
1993-94 NAEP U.S. History Assessment
Interreader Reliability Ranges

Assessment	Number of Unique Items	Number of Items in Percentage Exact Agreement Range				
		50-59%	60-69%	70-79%	80-89%	90-100%
Grade 4 (Age 9)						
Short Constructed-Response Items	25	0	0	0	8	17
Extended Constructed-Response Items	7	0	0	0	3	4
Grade 8 (Age 13)						
Short Constructed-Response Items	37	0	0	0	17	20
Extended Constructed-Response Items	10	0	0	0	7	3
Grade 12 (Age 17)						
Short Constructed-Response Items	33	0	0	1	17	15
Extended Constructed-Response Items	19	0	0	1	13	5

7.3.5 Main Geography Assessment

7.3.5.1 Characteristics of the Scoring Guides

The geography portion of the 1994 assessment included a total of 89 discrete constructed-response items. A variety of constructed-response items were utilized in the assessment in order to measure different elements of students' conceptual understanding of scientific material as well as practical reasoning ability. The items scored included short answer constructed-responses, extended constructed-responses, maps, diagrams, and graphs. Each constructed-response item had a unique scoring guide that identified the range of possible scores for the item and defined the criteria to be used in evaluating student responses. Due to the nature of the items and difficulty distinguishing between complete/correct responses and partially correct responses, a small number of items were scored dichotomously on a right/wrong basis.

The readers scoring the constructed-response items were organized into two teams. Team 1 consisted of eight readers who scored fourth- and eighth-grade items. During the course of the project they scored 34 short constructed-response items using a scale that allowed for partial credit as follows:

1 = Incorrect response
2 = Partial understanding
3 = Correct response

The readers also scored 10 extended constructed-response items on a scale of 1 to 4 as follows:

1 = Incorrect response
2 = Minimal understanding
3 = Satisfactory level of comprehension
4 = Extended response with correct reasoning

Four dichotomous items were also scored by Team 1 during the project. In these cases the items were quite straight forward and were scored on a scale of 1 and 3; a 1 being incorrect and 3 being correct with no score point of 2 being awarded. Additionally, scores for all items included a 0 for no response, an 8 for a completely erased or scratched-out response, and a 9 for any response found to be unratable i.e., illegible, completely off-task, or a response of "I don't know."

The seven readers on Team 2 concentrated on eighth- and twelfth-grade items and scored 25 short constructed-response items as well as 16 extended constructed-response items. As with Team 1, the scorers used a 1 to 3 or 1 to 4 scale for the items. In contrast to Team 1, there were no dichotomously scored items assigned to Team 2. The assignment of readers between the two teams was made on the basis of each reader's academic qualifications, teaching experience, and other scoring-related experiences.

7.3.5.2 Training

The training on each item was conducted by ETS geography specialists, with assistance from the two table leaders and the scoring specialist. For each team, the first training session began on March 28, 1994, with scoring of those items following immediately. After scoring all available responses, training on the next item proceeded. Although each team worked at a slightly different pace due in part to the levels of difficulty of their respective items, each team typically scored at least one item per day. Scoring was completed on May 23.

7.3.5.3 Scoring

Once the practice session was completed, the formal scoring process began. During the scoring, the table leaders continued to compile notes on scoring decisions for the readers' reference and guidance and for the permanent scoring records. Additionally, the table leaders closely monitored interreader reliability using both team and individual statistics as a reference, and performed back-reading duties in which they reviewed a sample of the papers scored by each reader on the team.

Because of the rapid pace of scoring, all responses to an item could typically be scored in a single session. Occasionally, however, scoring would extend over the lunch break or need to be completed the next morning. Whenever the readers on a team broke for more than 15 minutes, the team members 'recalibrated' by individually scoring a five-response calibration set. After the table leader reviewed the calibration set scores and was satisfied with the performance of the readers, scoring could resume. Additionally, the scoring specialist continuously monitored the progress of each team and noted all scoring-related decisions to ensure that training and scoring progressed smoothly and in a timely manner.

One innovative feature of geography scoring was the inclusion of three items in which the student used colored pencils to construct answers. Scoring of these items was organized in such a fashion that readers were trained on all three items at one time and then allowed to score the three items as a set.

One of the advantages of utilizing an image-based scoring system is the ability to construct reader aids to simplify scoring and thus increase reader reliability. Five items were identified by the ETS subject area specialist and the NCS scoring specialist as excellent candidates for the construction of 'overlays' that were displayed along with the student responses and served as templates that defined boundaries in which a correct response must be located. Schematic representations of these overlays are included with the scoring guides and sample papers for those five items.

A minimum of 25 percent of the geography booklets was scored by a second reader to obtain statistics on interreader reliability. These statistics can be found in summary in Table 7-7.

Table 7-7
1993-94 NAEP Geography Assessment Interreader Reliability Ranges

Assessment	Number of Unique Items	Number of Items in Percentage Agreement Range				
		50-59%	60-69%	70-79%	80-89%	90-100%
Grade 4 (Age 9)						
Short Constructed-Response Items	24	0	0	0	2	22
Extended Constructed-Response Items	7	0	0	0	1	6
Grade 8 (Age 13)						
Short Constructed-Response Items	33	0	0	0	5	28
Extended Constructed-Response Items	8	0	0	1	3	4
Grade 12 (Age 17)						
Short Constructed-Response Items	24	0	0	0	7	17
Extended Constructed-Response Items	14	0	0	1	10	3

7.3.6 Main Reading Assessment

7.3.6.1 Characteristics of the Scoring Guides

The 1994 main NAEP reading assessment for grades 4, 8, and 12 contained a total of 175 discrete constructed-response items. Of these, 93 short items were scored on a 2-point scale, 49 short items were scored on a 3-point scale, and 33 extended items were scored on a 4-point scale. These items were found in 25 discrete blocks accompanying specific reading passages. Of these 25 blocks, two blocks appeared in both grade 4 and grade 8 booklets and four blocks appeared in both grade 8 and grade 12 booklets. All overlap booklets contained the same passage at both grade levels. Except for one block (Q12R7 versions A & B) all overlap blocks contained the same constructed response items presented in the same order at both grade levels. Two cognitive blocks appeared in each student booklet, and each 25-minute block appeared in four different booklets at each appropriate grade level. The 50-minute blocks each appeared in only one booklet. There were a total of 16 grade 4 booklets, 21 grade 8 booklets, and 22 grade 12 booklets. Included in this distribution of blocks and items were the NAEP Reader blocks, for grades 8 and 12.

7.3.6.2 Training

The reading assessment followed the basic training procedures outlined above. One trainer provided all the training for the fourth-grade blocks scored on the evening shift. On the day shift, there

BEST COPY AVAILABLE

were usually four trainers working among the seven teams at any given time. One trainer followed the fourth-grade blocks from beginning to end. Another followed the NAEP reader items throughout the duration of the project. The eighth- and twelfth-grade blocks were split among various trainers on a weekly basis.

7.3.6.3 Scoring

Each constructed-response item had a unique scoring standard that identified the range of possible scores for the item and defined the criteria to be used in evaluating the students' responses. Point values were assigned with the following meanings:

2-point (Dichotomous) items from the 1992 assessment:
1 = Unacceptable
4 = Acceptable

2-point (Dichotomous) items developed during the 1993 field test:
1 = Evidence of little or no comprehension
3 = Evidence of full comprehension

3-point items developed during the 1993 field test:
1 = Evidence of little or no comprehension
2 = Evidence of partial or surface comprehension
3 = Evidence of full comprehension

All 4-point items:
1 = Evidence of unsatisfactory comprehension
2 = Evidence of partial comprehension
3 = Evidence of essential comprehension
4 = Evidence of extensive comprehension

The scores for these items also included a 0 for no response, an 8 for an erased or crossed-out response, and a 9 for any response found to be unratable (i.e., illegible, off-task, responses written in a language other than English, or responses of "I don't know"). The tax form also used the score point 6 to indicate that a completed tax form was not returned with the assessment booklet.

During scoring, the table leaders compiled notes on various responses for the readers' reference and guidance and for the permanent record. In addition, trainers were accessible for consultation in interpreting the guides for unusual or unanticipated responses. The table leaders conducted constant on-line back-reading of all team members' work throughout the scoring process, bringing to the attention of each reader any problems relating to scoring. When deemed appropriate, scoring issues were discussed among the team as a whole. Table leaders also monitored n-counts of responses scored and individual and team reliability figures throughout the course of scoring.

Each item was scored by a single team immediately after training for that item. Team sizes ranged from 7 to 13 scorers. Altogether, 17 teams scored the reading items—7 teams on the day shift and 10 teams on a shorter evening shift. Teams scored anywhere from 3 to 37 unique items during the course of the project.

Grade 4 items came from both national and state-by-state samples. Responses were delivered by image in such a way that the student demographics were unknown to the reader. Thus, readers did not know from which sample any given item came when it appeared on the screen. In the case of overlap items, all readers scored responses at both grade levels.

A minimum of 25 percent of the 1994 reading responses were scored a second time. The image system presented all responses in the same manner, so the reader could not discern which responses were being scored for the first time and which were being scored for the second time. The table leader and the ETS trainer were able to monitor these figures on demand. The system showed the overall reliability for the group scoring the item and individual reliability of the qualified readers.

7.3.6.4 NAEP Reader

The 1994 main NAEP reading assessment for grades 8 and 12 included two 50-minute blocks at each grade level for the *NAEP Reader*. This special study portion of the reading assessment consisted of two *NAEP Reader* booklets, one for each grade, each of which contained seven different stories. At each grade level, some students were allowed to choose a story and answer a series of questions about the story, while other students were assigned a story to read. All students were asked the same questions regardless of which story they read or whether they were allowed to choose or were assigned a story. The only exception was that the students who were allowed to choose a story were asked why they chose that story, while those who were assigned a story were not asked that question. Training and scoring followed the same procedures as the rest of the reading assessment, except that within each item, the responses were divided according to which story the student read. For each item scored, the readers went through seven training sets, one for each story. Thus, the team scored all responses for one item for one story before training for the same item for the next story. The reliability figures for the main reading assessment as well as the NAEP Reader are presented in Table 7-8.

Table 7-8

1993-94 NAEP Reading Assessment Interreader Reliability Ranges

Assessment	Number of Unique Items	Number of Items in Percentage Agreement Range			
		60-69%	70-79%	80-89%	90-100%
Reading					
Grade 4 (Age 9)					
Short Constructed-Response Items	37	0	0	8	29
Extended Constructed-Response Items	8	0	1	6	1
Grade 8 (Age 13)					
Short Constructed-Response Items	56	0	0	18	38
Extended Constructed-Response Items	13	0	1	9	3
Grade 12 (Age 17)					
Short Constructed-Response Items	63	0	0	30	33
Extended Constructed-Response Items	13	0	2	6	5
NAEP Reader					
Grade 8 (Age 13)					
Short Constructed-Response Items	9	0	0	6	3
Extended Constructed-Response Items	3	0	1	1	1
Grade 12 (Age 17)					
Short Constructed-Response Items	9	0	0	4	5
Extended Constructed-Response Items	3	0	0	2	1

7.3.6.5 1992 Short-Term Trend and Image/Paper Special Study

Sixteen blocks from the 1992 reading assessment were re-used in the 1994 assessment to provide data with which to study trends over time. To accomplish this, a random sample of responses from the 1992 assessment were pulled from the warehouse for rescoring to determine whether or not the scoring performed in 1994 was comparable to the scoring performed in 1992. For the national sample, ETS measurement personnel identified three booklets at grade 4, four booklets at grade 8, and five booklets at grade 12 that contained all of the blocks needed for the study. See Table 7-9 for a list of booklets chosen. The entire sample of those booklets was used for the rescore study. Since each block appears in four booklets, rescoring the entire sample of one booklet resulted in a 25 percent rescore of the responses from 1992. These booklets were scanned to capture the same clip areas used for the 1994 responses. Thus, they appeared identical to the reader when viewing them on the monitor and were presented at the same time as the 1994 responses.

Table 7-9
1993-94 NAEP Assessment
1992 National Reading Booklets and Blocks Rescored on Paper-Based and Image Systems

Year	Grade	Booklet	Block Number	Title	Block Number	Title
1992	4	R30	1R4	Box in Barn	1R3	Spider and Turtle
		R37	1R10	Wombats	12R7A	Ellis Island Grade 4
		R44	12R5	Money Makes Cares	1R6	Blue Crabs
	8	R32	12R5	Money Makes Cares	23R3	Flying Machine
		R35	12R7B	Ellis Island Grade 8	23R6	Gift of Phan
		R38	2R11	Time Capsule	23R10	Write Your Senator
		R48	2R12	NAEP Reader		
	12	R30	23R3	Flying Machine	3R4	Mountain Trail
		R33	23R6	Gift of Phan	3R7	Garbage Glut
		R38	3R11	Tax Form	23R10	Write Your Senator
		R48	3R12	NAEP Reader		
		R49	3R13	Call Me Gentle		

After scanning was completed, the national sample of the rescore booklets was transported to the scoring facilities to be scored on paper. Paper scoring took place at the same time as image scoring. From these studies, data were available to compare the paper-based scoring done in 1992, the paper-based scoring done in 1994, and the image-based scoring done in 1994. The results of this comparison suggest that the use of image-based scoring technology produces ratings equivalent to those from the traditional paper-based method (Johnson, 1993).

7.3.7 Calibration Bridges

Unanticipated delays in receipt and processing of student booklets resulted in a situation in which scoring for some constructed-response items began before all or most of the student responses for those items were available for scoring. The result was that the responses for most of the 1994 NAEP constructed-response items were scored in two different scoring sessions (sweeps). To maintain the highest standards of scoring and measurement precision and to ensure that calibration error was not introduced as a result of the split scoring sessions and the time elapsed between them, a plan was devised to calibrate the scoring of sweeps 1 and 2. In some instances, it was determined that scoring could resume with a review of training and a regular calibration set to ensure consistency and reliability. In other instances, a calibration bridge was constructed to provide statistical linkage between the two scoring sessions.

It was determined that scoring could continue without the calibration bridge in those instances in which completed scoring had met both of the following criteria:

- 50 percent scored on first sweep
- Interreader reliability equal to or greater than 95 percent

For those items not meeting the above criteria, a set of papers was scored to provide a reliability link or calibration bridge between completed scoring (first sweep) and subsequent scoring (second sweep). The procedures for completing the calibration bridge were as follows:

1. Approximately 12,500 processed booklets were pulled from inventory and from them samples of student responses were constructed for each item designated for the calibration bridge scoring.

2. A file of all pulled booklet ID numbers was created along with all scores assigned in the first sweep of scoring. This allowed for matching scores assigned in the first sweep to those given in the paper-based calibration bridge rescoring.

3. For each designated item, each scorer read and scored at least 10 student responses drawn from this sample (10 different papers for each scorer). No papers with scores of 0 were included, and 20 percent of the responses were scored twice for interreader reliability.

4. The clerical support staff entered the scores in a spreadsheet program that produced data on reader agreement, score distributions, mean scores, and standard deviations of the mean scores.

5. The data from the calibration bridge scoring was compared to the data for the first sweep scoring on the same item.

6. After reviewing this data, items meeting the following criteria were determined to be ones for which second sweep scoring could then proceed:

   a. items for which the Diff T test was not significant (i.e., the p-value for the statistic was >.05, and hence the null hypothesis cannot be rejected).

   b. items for which the calibration/bridge/sweep percent agreement was higher than the designated threshold.

2-point item	90% reliability
3-point item	80% reliability
4-point item	75% reliability

   c. items for which the calibration/bridge interreader reliability was no more than six percentage points lower than first sweep interreader reliability.

7. For those items not meeting the criteria, readers were retrained. Following the retraining, five different papers from the sample were read by each reader and results evaluated.

8. Following analysis of the results of scoring after this retraining, a decision was made to continue scoring or, alternately, to rescore all the previously scored responses along the remaining responses for the item under consideration.

Forty-three calibration bridges were conducted in U.S. history, 28 were conducted in geography, and 95 were conducted in reading. As a result of the calibration bridge, five items were determined to require complete rescoring.

Chapter 8

CREATION OF THE DATABASE, QUALITY CONTROL OF DATA ENTRY, AND CREATION OF THE DATABASE PRODUCTS[1]

John J. Ferris, David S. Freund, Patricia E. O'Reilly,
Katharine E. Pashley, and Alfred M. Rogers
Educational Testing Service

8.1 INTRODUCTION

The data transcription processing of the 1994 NAEP data was conducted by the ETS subcontractor, National Computer Systems (NCS). The processes conducted by NCS, described in Chapter 6, resulted in the transfer to ETS of data files containing response data for students, teachers, and school administrators. At the same time, NCS transferred to ETS subcontractor Westat, Inc., the demographic data needed to derive sampling weights. Westat then provided ETS with data files that included sampling weights for assessed students, excluded students, and schools, data, school characteristic data, and community characteristic data. Before any analysis could begin, these data files had to be merged into a comprehensive, integrated database. The creation of the database is described in Section 8.2.

To evaluate the effectiveness of the quality control of the data entry process, the final database was sampled and verified in detail against the original instruments received from the field. The results of this procedure are given in Section 8.3.

The integrated database was the source for the creation of two NAEP database products: the item information database and the secondary-use data files. These are described in Section 8.4.

8.2 CREATION OF THE DATABASE

8.2.1 Merging Files

The transcription process resulted in the generation of 12 sets of data files (long-term trend data and reading, geography, and U.S. history cross-sectional data for each of the three student cohorts). Included in each set were student response data, school questionnaire data, and excluded student questionnaire data. A fourth data file contained teacher questionnaire data for the teachers of fourth-grade and eighth-grade students assessed in reading, U.S. history or geography. The process of deriving sampling weights produced an additional three files (assessed student weights, excluded student weights, and school weights) for each of the 12 sets. Before data analyses could be performed, these files had to be integrated into a coherent and comprehensive database.

---

[1] John J. Ferris was responsible for the evaluation of the quality of the database and the data entry process; Katharine E. Pashley was responsible for database generation under the supervision of David S. Freund; Patricia E. O'Reilly and Alfred M. Rogers created the secondary-use data files.

The database ultimately comprised six files per cohort: four assessed student files (reading, geography, U.S. history, and long-term trend), a single school file, and a single excluded student file. The student files were separated by subject area to improve maintenance and efficiency of the databases and data analyses. The student data were created in several steps, merging the student response data with the demographic and community variables, the student weights, and key derived variables such as parental education and home environment composites. Also, teacher data were appended to student records in cases where the student's teacher responded to a teacher questionnaire. In all steps, the booklet serial number was used as the matching criterion.

The school file was created by matching and merging the school questionnaire file with the school weights file and with a file of school variables supplied by Westat, which included demographic information about the school and the community. The PSU and school code were used as the matching criteria. Each record of the resulting file was formed by merging the weight information with the response data and the school demographic and community data. Since not all schools returned their questionnaires, some of the output records contained only school identifying information and weight information. The school file could be accessed on its own or it could be linked to the assessed and excluded student files through the PSU and school codes.

The excluded student file was the result of merging the excluded student questionnaire file with the excluded student weights file. The booklet serial number was used as the matching criterion.

To ensure that the data were transferred accurately from NCS to ETS and that the processing control parameters used by analysis programs at ETS were properly in place, several quality control procedures were implemented. The principal procedure included matching independently generated frequency distributions computed at NCS with those produced at ETS, using ETS control parameters to process the data. Distributions for all student response questions (approximately 1,500 items) were verified to match.

When the appropriate files had been merged and proper quality control procedures had been completed, the database was ready for analysis. Any time that new data values, such as plausible values, were derived external to the database, they were added to the relevant files using the matching procedures described above. The secondary-use data files were later generated from this database.

8.2.2 Creating the Master Catalog

A critical part of any database is its processing control and descriptive information. A central repository of this information may be accessed by all analysis and reporting programs to provide correct parameters for processing the data fields as well as to provide consistent labeling to identify the results of the analyses. The NAEP master catalog file was designed and constructed to serve both of these purposes.

Each record of the master catalog contains the processing, labeling, classification, and location information for each data field in the NAEP database. The control parameters are used by the access routines in the analysis programs to define the manner in which the data values are to be transformed and processed.

Each data field has a 50-character label in the master catalog describing the contents of and, where applicable, the source of the field. The data fields with discrete or categorical values (e.g., multiple-choice items and constructed-response items, but not weight fields) have additional label fields in the catalog containing 8- and 20-character labels for those values.

The classification area of the catalog record contains distinct fields corresponding to predefined classification categories for the data fields. For a given classification field, a nonblank value indicates the code within that classification category for the data field. This permits the collection of identically classified items or data fields by performing a selection process on one or more classification fields in the catalog.

According to the NAEP design, it is possible for item data fields to appear in more than one student sample and in more than one block within each sample. The location fields of the catalog record contain the age, block and, where applicable, the sequence within the block for each appearance of the data field. (Fields such as plausible values and weights do not contain sequence numbers since these fields are not pertinent to a given block.)

The master catalog file was constructed in parallel with the collection and transcription of the assessment data to be ready for use by analysis programs when the database was created. As new data fields were derived and added to the database, their corresponding descriptive and control information were entered into the catalog.

One of the most important uses of the master catalog was the control of the creation of the secondary-use data files, codebooks, and file layouts. A synopsis of this process is presented in Section 8.4.

8.3 QUALITY CONTROL OF NAEP DATA ENTRY FOR 1994

This section describes the evaluation of the data entry process for the 1994 national assessment. As in past years, the NAEP database was found to be more than accurate enough to support the analyses that were done. Overall, the observed error rates were comparable to those of past assessments; they ranged from four errors per 10,000 responses for the student data to 18 errors per 10,000 responses for the IEP/LEP student questionnaire data.

The purpose of the analysis reported in this section is to assess the quality of the data resulting from the complete data entry system, beginning with the actual instruments collected in the field and ending with the final machine-readable database used in the analyses. The process involved the selection of instruments at random from among those returned from the field and the comparison of these instruments, character by character, with their representations in the final database. In this way, we were able to measure the error rates in the data as well as the success of the data entry system.

Of course the observed error rate cannot be taken at face value. For example, the sample of teacher questionnaires that happened to be selected for close inspection contained only one error out of a total of 3,515 characters. To conclude that the entire teacher questionnaire database has an error rate of 1/3515, or .0003, would be an act of extreme optimism; we may simply have been lucky with this particular random sample. What is needed is an indication of how bad the true error rate might be, given what we observed. Such an indication is provided by confidence limits. Confidence limits indicate how likely it is that a value falls inside a specified range of values in a specified context or distribution. In our analysis, the specified range is an error rate between zero and some maximum value beyond which we are confident at some specified level (99.8 percent) that the true error rate does not lie; the specified context or distribution turns out to be the cumulative binomial probability distribution. An example should demonstrate this technique:

> Let us say that 1,000 booklets were processed, each with 100 characters of data transcribed for a total of 100,000 characters. Let us say further that five of these characters were discovered to be in error in a random sample of 50 booklets that were completely checked; in other words, five errors were found in a sample of 5,000 characters. The following expression may be used to establish the probability that the true error rate is .0025 or less, rather than the single-value estimate of the observed rate of one in a thousand (.001):
>
> $$\sum_{j=0}^{5} \binom{5000}{j} \times .0025^{j} \times (1 - .0025)^{(5000-j)} = .0147$$
>
> This is the sum of the probability of finding five errors plus the probability of finding four errors plus. . . etc. . . plus the probability of finding zero errors in a sample of 5,000 with a true error rate of .0025; that is, the probability of finding five or fewer errors by chance when the true error rate is .0025. Notice that we did not use the size of the database in this expression. Actually, the assumption here is that our sample of 5,000 was drawn from a database that is infinite. The smaller the actual database is, the more confidence we can have in the observed error rate; for example, had there been only 5,000 in the total database, our sample would have included all the data, and the observed error rate would have been the true error rate. The result of the above computation allows us to say, conservatively, that .0025 is an upper limit on the true error rate with 98.53 percent (i.e., 1 - .0147) confidence; that is, we can be quite sure that our true error rate is no larger than .0025.

Virtually all of the data collected for this assessment were machine-scanned. The only exception was a set of six booklets used for the long-term trend reading and writing assessments; the format of these booklets was kept the same for comparability with earlier assessments, so these booklets had to be key-entered. As it happened, no errors at all were found in the sample of key-entered booklets selected for quality control.

In the 1994 assessment, the selection of booklets for this comparison took place at the point of first entry into the recording process for data from the field. In past assessments, this selection took place only after data had reached the final database, in order to assure that only relevant booklets were involved in the quality control evaluation. While the new method of selection did result in some irrelevant booklets (due to absent students or other problems) sufficient numbers of booklets that did appear in the final database were ultimately selected. The earlier availability of booklets for quality control evaluation and the improved efficiency of this new selection process were adequate compensation for the loss of control over which booklets were involved in quality control evaluation.

The individual instruments are briefly discussed in the following sections and a summary table (Table 8-1) gives the upper 99.8 percent confidence limit for the error rate for each of the instruments as well as the sampling information. The 99.8 percent confidence limit, and the selection rates noted, were chosen to make these results comparable to those of previous administrations when the same parameters were used.

Table 8-1

Summary of Quality Control Error Analysis for NAEP 1994 Data Entry

Instrument/Sample	Selection Rate	Different Booklets	Number of Booklets Sampled	Number of Characters Sampled	Number of Errors	Observed Error Rate	Upper 99.8% Confidence Limit
Student Booklets - Main	1/375	199	279	21,525	9	.0004	.0010
Student Booklets - Long-Term Trend	1/430	26	72	8,942	6	.0007	.0019
IEP/LEP Student Questionnaires	1/193	2	75	4,964	9	.0018	.0043
Teacher Questionnaires	1/120	3	24	3,515	1	.0003	.0024
School Characteristics and Policies Questionnaires	1/85	3	24	2,791	2	.0007	.0038

BEST COPY AVAILABLE

212

211

8.3.1 Student Data

For the first time, the data entry process relied on image processing technology for recording the scores assigned by professional readers to the students' responses to constructed-response items. The scanned image of a student's response to one of these items was presented on the computer screen of a reader's work station. After determining the score for the item, the reader then entered this score using the keyboard at the work station.

This new process raised the question of what to verify or check in a quality control operation. The usual issue—whether the response that ended up in the final database is the same as the original intended response—could not be raised here, since the reader's intention, which defines the data, was entered *directly into the database* without any intermediate steps. The question of whether readers consistently and accurately applied agreed-upon scoring rubrics was not at issue here; that question falls into the province of reader reliability studies. In short, the data for these items existed in only one form, the database itself, and could not be verified against any earlier or preliminary form.

Rather than abdicate all quality control responsibility for these items, we chose to verify the process itself. Two important questions were examined:

1. Was the identity of the respondent maintained? Did a respondent's scores end up in his or her data record and not someone else's?
2. Was the identity of the item maintained? Did the score for each constructed-response item in this booklet end up correctly identified in the database, or was it transposed with another item response or perhaps left out?

There were a total of 45 different booklets in this assessment that contained some number of constructed-response items requiring professional scoring. To verify that the system was functioning correctly, 45 sets of artificial data were carefully constructed, one set for the constructed-response items in each of these booklets. Each set consisted of two booklets, representing altogether a total of 90 "respondents." These 90 booklets were filled in with preassigned scores and processed in the usual way, the only difference being that the readers were presented with the score to assign, rather than with a passage to be evaluated.

To assure correct identification of a booklet (question #1 above), the score pattern of each booklet was made unique, even under the assumption that the scorers made one recording error in every booklet. Such an error would not be relevant to the question of whether the correct respondent was being scored. As noted above, the question of whether the correct score is being assigned needs to be addressed through reader reliability studies.

To assure that item identity was maintained within a booklet (question #2 above), different responses were used across the constructed-response items for each booklet. Since the number of different responses was almost never adequate to allow making each response unique within a booklet, a second sample of each booklet was needed. Any item response which had to be duplicated within the first booklet of such a pair was designed to be different in the second booklet, and vice versa.

Both respondent identity and item identity were maintained satisfactorily. While the correct artificial data did not always turn up in the database, the discrepancies were almost entirely due to a failure to properly communicate the artificial data to the readers. The placement of the quality control item labels (described in Chapter 6) was especially critical for geography items that involved map-reading. A number of errors occurred here that would not have occurred had the readers been dealing with real data rather than artificial data; such errors were not considered relevant to assessing the quality of the real database. It is difficult to design the input of artificial data in such a way that it will pass for real data without introducing extraneous and irrelevant errors; perhaps an alternative process should be developed to accomplish this. Notwithstanding these problems, this part of the quality control effort was successful enough to demonstrate that both item and respondent integrity were maintained in these booklets of artificial data.

About 135,000 students were assessed across all samples in this assessment. Across all student data, roughly one booklet in 385 was selected for close examination, which is comparable to the one in 400 target selection rate used in past assessments. The student data error rates, while slightly higher than in recent assessments, were consistently low in all subject areas and across all three grades. The overall quality of the data was very high.

8.3.2 IEP/LEP Student Questionnaire Data

In this assessment, 14,496 questionnaires were scanned. Most of these were IEP/LEP questionnaires; 2,294 of the older excluded student questionnaires were used in connection with the long-term trend assessments for continuity with earlier years, but they differed from the newer IEP/LEP questionnaires by only one item. The quality control sampling rate was one in 193, comparable to the rate used in previous assessments. The results were also much like the results from previous assessments—a higher error rate than that for other instruments, but not alarmingly so. All of the problems encountered involved the scanner's mistaking an erasure for a genuine response.

8.3.3 Teacher Questionnaire Data

In this assessment, 2,892 teacher questionnaires were collected and scanned. About one percent of these questionnaires was sampled for the quality control procedure. The error rates for these questionnaires were somewhat higher than those of the student data, and somewhat lower than those found in past assessments. The error found involved the failure of the scanner to pick up a legitimate mark.

8.3.4 School Characteristics and Policies Questionnaire Data

In this assessment, 2,039 school characteristics and policies questionnaires were collected. They were sampled at a rate of about one in 85. Two scanning errors were found in the questionnaires that were checked.

8.4 NAEP DATABASE PRODUCTS

The NAEP database described to this point serves primarily to support analysis and reporting activities that are directly related to the NAEP contract. This database has a singular structure and access methodology that is integrated with the NAEP analysis and reporting programs. One of the directives of the NAEP contract is to provide secondary researchers with a nonproprietary version of the database that is portable to any computer system. In the event of transfer of NAEP to another client, the contract further requires ETS to provide a full copy of the internal database in a format that may be installed on a different computer system.

In fulfillment of these requirements, ETS provides two sets of database products: the item information database and the secondary-use data files. The contents, format and usage of these products are documented in the publications listed under the appropriate sections below.

8.4.1 The Item Information Database

The NAEP item information database contains all of the descriptive, processing, and usage information for every assessment item developed and used for NAEP since 1970. The primary unit of this database is the item. Each NAEP item is associated with different levels of information, including usage across years and age cohorts, subject area classifications, response category descriptors, and locations of response data on secondary-use data files.

The item information database is used for a variety of essential NAEP tasks: providing statistical information to aid in test construction, determining the usage of items across assessment years and ages for trend and cross-sectional analyses, labeling summary analyses and reports, and organizing items by subject area classifications for scaling analysis.

The creation, structure, and use of the NAEP item information database for all items used up to and including the 1994 assessment are fully documented in the NAEP publications *A Guide to the NAEP Item Information Database* (Rogers, Barone, & Kline, 1996) and *A Primer for the NAEP Item Information Database* (Rogers, Kline, Barone, Mychajlowycz, & Forer, 1989).

The procedures used to create the 1994 version of the item information database are the same as those documented in the guide. The version of the guide contains the subject area classification categories for the cognitive items.

8.4.2 The Secondary-Use Data Files

The secondary-use data files are designed to enable any researcher with an interest in the NAEP database to perform secondary analysis on the same data as those used at ETS. The three elements of the distribution package are the data files, the printed documentation, and copies of the questionnaires and released item blocks. A set of files for each sample or instrument contains the response data file, a file of control statements that will generate an SPSS system file, a file of control statements that will generate a SAS system file, and a machine-readable catalog file. Each machine-readable catalog file contains sufficient control and descriptive information to permit the user who does not have either SAS or SPSS to set up and perform data analysis. The printed documentation consists of four volumes: a guide to the use of the data files, and a set of data file layouts and codebooks for each of the three age cohorts (see the *National Assessment of Educational Progress 1994 Trial State Assessment Program in Reading Secondary-Use Data Files User Guide*, O'Reilly, Zelenak, Rogers, & Kline, 1996a).

The remainder of this section summarizes the procedures used in generating the data files and related materials.

8.4.2.1 File Definition

The design of the 1994 assessment perpetuates two features of the 1990 and 1992 assessment design: the focused-BIB booklet design and the direct matching of teacher questionnaires to student assessment instruments.

The focused-BIB design within the main assessment isolates the primary subject areas to separate groups of booklets. This permits the division of the main sample into subject-specific subsamples. The data files generated from these subsamples need only contain the data that is relevant to their corresponding subject areas and are therefore smaller and more manageable than their counterparts in previous assessments.

According to the design of the 1984, 1986, and 1988 assessments, only a sample of the teachers of the assessed students were asked to fill out the teacher questionnaires. The large size of the secondary-use main student files and the relatively low matching rate between students and teachers made it impractical if not physically prohibitive to produce a complete file with student and teacher information. Both the 1984 and 1986 secondary-use data packages had separate teacher data files which could be linked to the student data files for analysis. The teacher file in the 1988 secondary-use data package contained not only the teacher response data, but also the data from the students who could be matched to teacher questionnaires. This type of file was more appropriate for the analysis of teacher data because it defined the student as the unit of observation.

The intent of the 1994 assessment design was to collect data from the reading, U.S. history, or geography teachers of the main assessment students at specified age/grade levels who were administered reading, U.S. history, or geography booklets. A portion of the teacher questionnaire contained questions that were directly related to each matched student. This change in the design afforded a very high matching rate between student and teacher data. Therefore, for those subject areas in each age/grade cohort for which teacher data were collected, the teacher responses were appended to each student record in the secondary-use data files.

8.4.2.2 Definition of the Variables

The lifting of the restraint on confidential data simplified the variable definition process as it permitted the transfer of *all* variables from the data base to the secondary-use files.

The initial step in this process was the generation of a LABELS file of descriptors of the variables for each data file to be created. Each record in a LABELS file contains, for a single data field, the variable name, a short description of the variable, and processing control information to be used by later steps in the data generation process. This file could be edited for deletion of variables, modification of control parameters, or reordering of the variables within the file. The LABELS file is an intermediate file only; it is not included on the released data files.

The next program in the processing stream, GENLYT, produced a printed layout for each file from the information in its corresponding LABELS file. These layouts were initially reviewed for the ordering of the variables.

The variables on all data files are grouped and arranged in the following order: identification information, weights, derived variables, proficiency scores (where applicable), and response data. On the student data files, these fields are followed by the teacher response data and the IEP/LEP student questionnaire data, where applicable. The identification information is taken from the front covers of the instruments. The weight data include sample descriptors, selection probabilities, and replicate weights for the estimation of sampling error. The derived data include sample descriptions from other sources and variables that are derived from the response data for use in analysis or reporting.

For each subject area of the main assessment, the item response data within each block were left in their order of presentation. The blocks, however, were arranged according to the following scheme: common background, subject-related background, the cognitive blocks in ascending numerical order, and student motivation. The responses to cognitive blocks that were not present in a given booklet were left blank, signifying a condition of "missing by design."

In order to process and analyze the spiral sample data effectively, the user must also be able to determine, from a given booklet record, which blocks of item response data were present and their relative order in the instrument. This problem was remedied by the creation of a set of control variables, one for each block, which indicated not only the presence or absence of the block but its order in the instrument. These control variables are included with the derived variables.

8.4.2.3 Data Definition

To enable the data files to be processed on any computer system using any procedural or programming language, it was desirable that the data be expressed in numeric format. This was possible, but not without the adoption of certain conventions for reexpressing the data values.

During creation of the NAEP database, the responses to all multiple-choice items were transcribed and stored in the database using the letter codes printed in the instruments. This scheme afforded the advantage of saving storage space for items with 10 or more response options, but at the expense of translating these codes into their numeric equivalents for analysis purposes. The response data fields for most of these items would require a simple alphabetic-to-numeric conversion. However, the data fields for items with 10 or more response choices would require "expansion" before the conversion, since the numeric value would require two column positions. One of the processing control parameters on the LABELS file indicates whether or not the data field is to be expanded before conversion and output.

The ETS database contained special codes to indicate certain response conditions: "I don't know" responses, multiple responses, omitted responses, not-reached responses, and unresolvable responses, which include out-of-range responses and responses that were missing due to errors in printing or processing. The scoring guides for the U.S. history, geography, and reading constructed-response items included additional special codes for ratings of "illegible," "erased" or "crossed out," and for "I don't know," "off task," or non-rateable by the scorers. All of these codes had to be reexpressed in a consistent numeric format.

The following convention was adopted and used in the designation of these codes: The "I don't know" and non-rateable response codes were always converted to 7; the "omitted" response codes were converted to 8; the "not-reached" response codes were converted to 9; the multiple response codes were converted to 0; and the "illegible," "erased," and "crossed out" response codes were converted to 5. The out-of-range and missing responses were coded as blank fields, corresponding to the "missing by design" designation.

This coding scheme created conflicts for those multiple-choice items that had seven or more valid response options as well as the "I don't know" response and for those constructed-response items whose scoring guide had five or more categories. These data fields were also expanded to accommodate the valid response values and the special codes. In these cases, the special codes were "extended" to fill the output data field: the "I don't know" and non-rateable codes were extended from 7 to 77, omitted response codes from 8 to 88, etc.

Each numeric variable on the secondary-use files was classified as either continuous or discrete. The continuous variables include the weights, proficiency scores, identification codes, and item responses where counts or percentages were requested. The discrete variables include those items for which each numeric value corresponds to a response category. The designation of "discrete" also includes those derived variables to which numeric classification categories have been assigned. The constructed-response items were treated as a special subset of the discrete variables and were assigned to a separate category to facilitate their identification in the documentation.

8.4.2.4 Data File Layouts

The data file layouts, as mentioned above, were the first user product to be generated in the secondary-use data files process. The generation program, GENLYT, used a LABELS file and a CATALOG file as input and produced a printable file. The LAYOUT file is little more than a formatted listing of the LABELS file.

Each line of the LAYOUT file contains the following information for a single data field: sequence number, field name, output column position, field width, number of decimal places, data type, value range, key or correct response value, and a short description of the field. The sequence number of each field is implied from its order on the LABELS file. The field name is an 8-character label for the field that is to be used consistently by all secondary-use data files materials to refer to that field on that file. The output column position is the relative location of the beginning of that field on each record for that file, using bytes or characters as the unit of measure. The field width indicates the number of columns used in representing the data values for a field. If the field contains continuous numeric data, the value under the number of decimal places entry indicates how many places to shift the decimal point before processing data values.

The data type category uses five codes to designate the nature of the data in the field: Continuous numeric data are coded "C"; discrete numeric data are coded "D"; constructed-response item data are coded "OS" if the item was dichotomized for scaling and "OE" if it was scaled under a polytomous response model. Additionally, the discrete numeric fields that include "I don't know" response codes are coded "DI." If the field type is discrete numeric, the value range is listed as the minimum and maximum permitted values separated by a hyphen to indicate range. If the field is a response to a scorable item, the correct option value, or key, is printed; if the field is an assigned score that was scaled as a dichotomous item using cut point scoring, the range of correct scores is printed. Each variable is further identified by a 50-character descriptor.

8.4.2.5 Data File Catalogs

The LABELS file contains sufficient descriptive information for generating a brief layout of the data file. However, to generate a complete codebook document, substantially more information about the data is required. The CATALOG file provides most of this information.

The CATALOG file is created by the GENCAT program from the LABELS file and the 1994 master catalog file. Each record on the LABELS file generates a CATALOG record by first retrieving the master catalog record corresponding to the field name. The master catalog record contains usage, classification, and response code information, along with positional information from the LABELS file: field sequence number, output column position, and field width. Like the LABELS file, the CATALOG file is an intermediate file and is not included on the released data files.

The information for the response codes, also referred to as "foils," consists of the valid data values for the discrete numeric fields, and a 20-character description of each. The GENCAT program uses additional control information from the LABELS file to determine if extra foils should be generated and saved with each CATALOG record. The first flag controls generation of the "I don't know" or non-rateable foil; the second flag regulates omitted or not-reached foil generation; and the third flag denotes the possibility of multiple responses for that field and sets up an appropriate foil. All of these control parameters, including the expansion flag, may be altered in the LABELS file by use of a text editor, in order to control the generation of data or descriptive information for any given field.

The LABELS file supplies control information for many of the subsequent secondary-use data processing steps. The CATALOG file provides detailed information for those and other steps.

8.4.2.6 Data Codebooks

The data codebook is a printed document containing complete descriptive information for each data field. Most of this information originates from the CATALOG file; the remaining data comes from the COUNTS file and the IRT parameters file.

Each data field receives at least one line of descriptive information in the codebook. If the data type is continuous numeric, no more information is given. If the variable is discrete numeric, the codebook lists the foil codes, foil labels, and frequencies of each value in the data file. Additionally, if the field represents an item used in IRT scaling, the codebook lists the parameters used by the scaling program.

Certain blocks of cognitive items in the 1994 assessment that are to be used again in later assessments for trend comparisons have been designated as nonreleased. In order to maintain their confidentiality, generic labels have been substituted for the response category descriptions of these items in the data codebooks and the secondary-use files.

The frequency counts are not available on the catalog file, but must be generated from the data. The GENFREQ program creates the COUNTS file using the field name to locate the variable in the database, and the foil values to validate the range of data values for each field. This program also serves as a check on the completeness of the foils in the CATALOG file, as it flags any data values not represented by a foil value and label.

The IRT parameter file is linked to the CATALOG file through the field name. Printing of the IRT parameters is governed by a control flag in the classification section of the CATALOG record. If an item has been scaled for use in deriving the proficiency estimates, the IRT parameters are listed to the right of the foil values and labels, and the score value for each response code is printed to the immediate right of the corresponding frequency.

The LAYOUT and CODEBOOK files are written by their respective generation programs to print-image disk data files. Draft copies are printed and distributed for review before the production copy is generated. The production copy is printed on an IBM 3800 printer that uses laser-imaging technology to produce high-quality, reproducible documentation.

8.4.2.7 Control Statement Files for Statistical Packages

An additional requirement of the NAEP cooperative agreement is to provide, for each secondary-use data file, a file of control statements each for the SAS and SPSS statistical systems that will convert the raw data file into the system data file for that package. Two separate programs, GENSAS and GENSPX, generate these control files using the CATALOG file as input.

Each of the control files contains separate sections for variable definition, variable labeling, missing value declaration, value labeling, and creation of scored variables from the cognitive items. The variable definition section describes the locations of the fields, by name, in the file, and, if applicable, the number of decimal places or type of data. The variable label identifies each field with a 50-character description. The missing value section identifies values of those variables that are to be treated as missing and excluded from analyses. The value labels correspond to the foils in the CATALOG file. The code values and their descriptors are listed for each discrete numeric variable. The scoring section is provided to permit the user to generate item score variables in addition to the item response variables.

Each of the code generation programs combines three steps into one complex procedure. As each CATALOG file record is read, it is broken into several component records according to the information to be used in each of the resultant sections. These record fragments are tagged with the field sequence number and a section sequence code. They are then sorted by section code and sequence number. Finally, the reorganized information is output in a structured format dictated by the syntax of the processing language.

The generation of the system files accomplishes the testing of these control statement files. The system files are saved for use in special analyses by NAEP staff. These control statement files are included on the distributed data files to permit users with access to SAS and/or SPSS to create their own system files.

8.4.2.8 Machine-Readable Catalog Files

For those NAEP data users who have neither SAS nor SPSS capabilities, yet require processing control information in a computer-readable format, the distribution files also contain machine-readable catalog files. Each machine-readable catalog record contains processing control information, IRT parameters, and foil codes and labels.

8.4.2.9 NAEP Data on Disk

The complete set of secondary-use data files described above are available on CD-ROM as part of the NAEP Data on Disk product suite. This medium can be ideal for researchers and policy makers operating in a personal computing environment.

The NAEP Data on Disk product suite includes two other components which facilitate the analysis of NAEP secondary-use data. The PC-based NAEP data extraction software, NAEPEX, enables users to create customized extracts of NAEP data and to generate SAS or SPSS control statements for preparing analyses or generating customized system files. The NAEP analysis modules, which currently run under SPSS® for Windows™, use output files from the extraction software to perform analyses that incorporate statistical procedures appropriate for the NAEP design.

Chapter 9

OVERVIEW OF PART II: THE ANALYSIS OF 1994 NAEP DATA[1]

Nancy L. Allen and James E. Carlson
Educational Testing Service

9.1 INTRODUCTION

This chapter presents an overview of the analyses conducted on the 1994 NAEP data, focusing on the common elements of the analyses used across the subject areas of the assessment. Details by subject area are provided in Chapters 12 through 18.

As described in Chapter 1, the 1994 NAEP comprised three major components. One component encompassed major assessments in reading, U.S. history, and geography, providing detailed information about student proficiency at the fourth-, eighth-, and twelfth-grade levels of nonpublic and public schools. Long-term trend assessments of reading, mathematics, science, and writing at ages 9, 13, and 17 constituted the second component. The third major component was the Trial State Assessment at the fourth-grade level in reading. Technical details of the Trial State Assessment are not included in this report but are presented in the *Technical Report of the NAEP 1994 Trial State Assessment Program in Reading* (Mazzeo, Allen, & Kline, 1995).

In addition to the three major components, a special study was conducted on choice of reading passage and its relationship with performance. This study involved the administration to eighth- and twelfth-grade students of a "NAEP Reader" that contained seven stories. Some students were directed to read a specific story, and others were instructed to select a story that interested them. The results from and procedures used in this special study are reported in a separate document. Likewise, results from a pilot study to examine the performance of a small group of students on a U.S. history task are reported in a separate document.

Results from the analyses described in the following chapters were reported in the following reports:

- *1994 NAEP Reading: A First Look—Findings from the National Assessment of Educational Progress* (Williams, Reese, Campbell, Mazzeo, & Phillips, 1995),

- *NAEP 1994 Reading Report Card for the Nation and the States: Findings from the National Assessment of Educational Progress and Trial State Assessments* (Campbell, Donahue, Reese, & Phillips, 1996),

---

[1] Nancy L. Allen was responsible for the psychometric and statistical analyses of national and state NAEP data. James E. Carlson was responsible for psychometric and statistical analyses relating to special aspects and issues of NAEP. Eugene G. Johnson, John Mazzeo, Spencer S. Swinton, and Rebecca Zwick also contributed to this chapter.

- *NAEP 1994 U.S. History: A First Look—Findings from the National Assessment of Educational Progress* (Williams, Lazer, Reese, & Carr, 1995),
- *NAEP 1994 U.S. History Report Card: Findings from the National Assessment of Educational Progress* (Beatty, Reese, Persky, & Carr, 1996),
- *NAEP 1994 Geography: A First Look—Findings from the National Assessment of Educational Progress* (Williams, Reese, Lazer, & Shakrani, 1995), and
- *NAEP 1994 Geography Report Card: Findings from the National Assessment of Educational Progress* (Persky, Reese, O'Sullivan, Lazer, Moore, & Shakrani, 1996).

9.2 SAMPLES OF STUDENTS

Because the samples of students included in the 1994 NAEP assessment are listed and described in detail in Chapter 1, only a brief description of these samples is given here. The 1994 national samples were of three general types: main NAEP samples, which were based on a common set of assessment procedures, including calendar-year age definitions; long-term trend samples, the purpose of which was to provide links to earlier assessments; and special study samples, used to examine the effect of choice of reading passage or reading proficiency and the U.S. history proficiency of students working in groups.

To shorten the timetable for reporting results, the period for national main assessment data collection was shortened in 1992 and 1994 from the five-month period (January through May) used in 1990 and earlier assessments to a three-month period in the winter (January through March, corresponding to the period used for the winter half-sample of the 1990 National Assessment).

The 1994 analyses of long-term trend data extended the trend lines commencing in 1971 in reading, 1973 in mathematics, 1969 in science, and 1984 in writing.

As described in Chapters 1 and 4, for each subject area in the main assessment, blocks of items were used to create a large number of different assessment booklets according to a focused balanced incomplete block (focused-BIB) design or according to a focused partially balanced incomplete block (focused-PBIB) design. The focused-BIB design provided for booklets that typically included three blocks of cognitive items in a single subject area, as well as background items. The blocks of cognitive items for reading, U.S. history, and geography included both multiple-choice and constructed-response items. With some variation, each block of cognitive items typically appeared in the same number of booklets. To balance possible block position main effects, each block appeared an equal number of times in each position. In addition, the BIB design required that each block of items be paired in a booklet with every other block of items exactly once. In the partially balanced design, each block of items appeared an equal number of times in each position, but was not necessarily paired in a booklet with every other block of items exactly once.

9.3 ANALYSIS STEPS

Because the analysis methods are not identical across subject areas or across major national and long-term trend samples, a separate analysis chapter has been included for each major assessment and for each long-term trend assessment. The procedures used depended on whether test items were scored dichotomously (right versus wrong) or polytomously (more than two categories of response) and whether links across age groups and/or across assessments were required. Basic procedures common to most or all of the subject area analyses are summarized here. The order is essentially that in which the procedures were carried out.

9.3.1 Preparation of Final Sampling Weights

Because NAEP uses a complex sampling design (Chapter 3) in which students in certain subpopulations have different probabilities of inclusion in the sample, the data collected from each student must be assigned a weight to be used in analyses. The 1994 NAEP weights were provided by Westat, Inc., the NAEP subcontractor in charge of sampling. Detailed information about the weighting procedures is available in Chapter 10 and in a report prepared by Westat personnel (Wallace & Rust, 1996).

9.3.2 Analysis of Item Properties: Background and Cognitive Items

The first step in the analysis of the 1994 data was item-level analysis of all instruments. Item analyses were performed separately for each grade level on each item in each subject area. Each block of items was analyzed separately, by grade, with the total score on the block (including the analyzed item) used as the criterion score for statistics requiring such a score. In the cases where final weights were not available, preliminary weights were used in these preliminary analyses. The item analysis of cognitive items was repeated after scaling of the items was completed. The results for only scaled items using final weights are reported in Chapters 12, 13, and 14.

➪ Background Items

For each NAEP background item, the unweighted and weighted percent of students who gave each response were examined, as well as the percent of students who omitted the item and the percent who did not reach the item. The number of respondents was also tabulated. These preliminary analyses were conducted within age/grade cohort and within major reporting categories. If unexpected results were found, the item data and the encoding of responses were rechecked.

➪ Cognitive Items

All NAEP cognitive items were subjected to analyses of item properties. These analyses included conventional item analyses and analyses of differential item functioning, and incorporated examinee sampling weights. Item analysis was conducted at the block level so that "number-correct" scores for an item are the total number of correct responses for the block containing that item. Because of the inclusion of polytomously scored items in the cognitive instruments, it was necessary to develop special procedures

BEST COPY AVAILABLE

for these items. The resulting statistics are analogous to those for the dichotomously scored items, as listed below.

Dichotomously Scored Items. These items were analyzed using standard procedures that result in a report for each item that includes:

- for each option of the item, for examinees omitting and not reaching the item, and for the total sample of examinees:
  - number of examinees,
  - percentage of examinees,
  - mean of number-correct scores, and
  - standard deviation of number-correct scores;
- p+, the proportion of examinees that received a correct score on the item (ratio of number correct to number correct plus wrong plus omitted);
- Δ, the inverse-normally transformed p+ scaled to mean 13 and standard deviation 4;
- the biserial correlation coefficient between the item and the number-correct scores; and
- the point-biserial correlation coefficient between the item and the number-correct scores.

Polytomously Scored Items. Enhanced procedures were employed for polytomously scored items. Methods parallel to those used for dichotomously scored items resulted in values reported for each distinct response category for the item. Response categories for each item were defined in two ways, one based on the original codes for responses as specified in the scoring rubrics used by the scorers and one based on a scoring guide developed by subject area and measurement experts. For example, a constructed-response item with four response categories would initially have six categories (not-reached, omitted, and the four valid response categories). These categories were used by the scorers in the scoring process. Another set of statistics resulted from mapping the response categories (excluding not-reached) into a new set of categories reflecting the scoring guide for the item. A constructed-response item with ordered categories, for example, would be mapped into a set of integers in a corresponding order. The scoring guide could result in the collapsing of (combining of) some response categories.

For these items the following statistics, analogous to those for dichotomously scored items, were computed:

- in place of p+, the ratio of the mean item score to the maximum-possible item score was used;

- in place of Δ, the ratio of the mean item score to the maximum-possible item score underwent the same transformation as that used on p+ to get Δ for dichotomously scored items;
- the polyserial correlation coefficient was used in place of the biserial; and
- the Pearson correlation coefficient was used in place of the point-biserial.

➪ Tables of Item-Level Results

Tables were created of the percentages of students choosing each of the possible responses to each item within each of the samples administered in 1994. The results for each item were cross-tabulated against the basic reporting variables such as region, gender, race/ethnicity, public/nonpublic school, and parental education. All percentages were computed using the sampling weights.

The sampling variability of all population estimates was obtained by the jackknife procedure used by ETS in previous assessments. Details of these procedures are presented in Chapter 10.

➪ Differential Item Functioning Analysis of Cognitive Items

Differential item functioning (DIF) analysis refers to procedures to assess whether items are differentially difficult for different groups of examinees. DIF procedures typically control for overall between-group differences on a criterion, usually by matching examinees from the two groups on overall test scores. Between-group performance on each item is then compared within sets of examinees having the same total test scores.

The DIF analyses of the dichotomous items were based on the Mantel-Haenszel chi-square procedure, as adapted by Holland and Thayer (1988). The procedure tests the statistical hypothesis that the odds of correctly answering an item are the same for two groups of examinees that have been matched on some measure of proficiency (usually referred to as the matching criterion). The DIF analyses of the polytomous items were based on the Mantel procedure (Mantel, 1963). These procedures compare proportions of matched examinees from each group in each polytomous item response category. The groups being compared are often referred to as the focal group (usually a minority or other group of interest, such as Black examinees or female examinees) and the reference group (usually White examinees or male examinees).

For both types of analyses, the measure of proficiency used is typically the total item score on some collection of items. Since, by the nature of the BIB design, booklets comprise different combinations of blocks, there is no single set of items common to all examinees. Therefore, for each student, the measure of proficiency used was the total item score on the entire booklet. These scores were then pooled across booklets for each analysis. This procedure is described by Allen and Donoghue (1994, in press). In addition, because research results (Zwick & Grima, 1991) strongly suggest that sampling weights should be used in conducting DIF analyses, the weights were used.

For each dichotomous item in the assessment, an estimate was produced of the Mantel-Haenszel common odds-ratio, expressed on the ETS delta scale for item difficulty. The estimates indicate the difference between reference group and focal group item difficulties (measured in ETS delta scale units), and typically run between about +3 and -3. Positive values indicate items that are differentially easier for the focal group than the reference group after making an adjustment for the overall level of proficiency in the two groups. Similarly, negative values indicate items that are differentially harder for the focal group than the reference group. It is common practice at ETS to categorize each item into one of three categories (Petersen, 1988): "A" (items exhibiting no DIF), "B" (items exhibiting a weak indication of DIF), or "C" (items exhibiting a strong indication of DIF). Items in category "A" have Mantel-Haenszel common odds ratios on the delta scale that do not differ significantly from 0 at the alpha = .05 level or are less than 1.0 in absolute value. Category "C" items are those with Mantel-Haenszel values that are significantly greater than 1 and larger than 1.5 in absolute magnitude. Other items are categorized as "B" items. A plus sign (+) indicates that items are differentially easier for the focal group; a minus sign (-) indicates that items are differentially more difficult for the focal group.

The ETS/NAEP DIF procedure for polytomous items uses the Mantel-Haenszel ordinal procedure. The summary tables of identified polytomous items contain generalizations of the dichotomous "A," "B," and "C" categories: "AA," "BB," and "CC."

Analyses were conducted for the main assessments in reading, U.S. history, and geography. Each set of analyses involved three reference group/focal group comparisons: male/female, White/Black, and White/Hispanic.

All analyses used rescaled sampling weights. A separate rescaled weight was defined for each comparison as

$$\textit{Rescaled Weight} = \textit{Original Weight} \times \frac{\textit{Total Sample Size}}{\textit{Sum of the Weights}}$$

where the total sample size is the total number of students for the two groups being analyzed (e.g., for the White/Hispanic comparison, the total number of White and Hispanic examinees in the sample at that grade), and the sum of the weights is the sum of the sampling weights of all the students in the sample for the two groups being analyzed. Three rescaled weights were computed for White examinees—one for the gender comparison and two for the race/ethnicity comparisons. Two rescaled overall weights were computed for the Black and Hispanic examinees—one for the gender comparison and another for the appropriate race/ethnicity comparison. The rescaled weights were used to ensure that the sum of the weights for each analysis equaled the number of students in that comparison, thus providing an accurate basis for significance testing.

In the calculation of total item scores for the matching criterion, both not-reached and omitted items were considered to be wrong responses. Polytomous items were weighted more heavily in the formation of the matching criterion, proportional to the number of score categories. For each item, calculation of the Mantel-Haenszel statistic did not include data from examinees who did not reach the item in question.

Each DIF analysis was a two-step process. In the initial phase, total item scores were formed, and the calculation of DIF indices was completed. Before the second phase, the matching criterion was refined by removing all "C" or "CC" items, if any, from the total item score. The revised score was used in the final calculation of all DIF indices. Note that when analyzing an item classified as "C" or "CC" in the initial phase, that item score is added back into the total score for the analysis of that item only.

Following standard practice at ETS for DIF analyses conducted on final forms, all "C" and "CC" items were reviewed by a committee of trained test developers and subject-matter specialists. Such committees are charged with making judgments about whether or not the differential difficulty of an item is *unfairly* related to group membership. The committee assembled to review NAEP items included both ETS staff and outside members with expertise in the field. The committee carefully examined each identified item to determine if either the language or contents would tend to make the item more difficult for an identified group of examinees. It was the committee's judgment that none of the "C" and "CC" items in the national assessment were functioning differentially due to factors irrelevant to test objectives. Hence, none of the items were removed from scaling due to differential item functioning. As pointed out by Zieky (1993):

> It is important to realize that *DIF* is not a synonym for *bias*. The item response theory based methods, as well as the Mantel-Haenszel and standardization methods of DIF detection, will identify questions that are not measuring the same dimension(s) as the bulk of the items in the matching criterion....Therefore, judgement is required to determine whether or not the difference in difficulty shown by a DIF index is *unfairly* related to group membership. The judgement of fairness is based on whether or not the difference in difficulty is believed to be related to the construct being measured....The fairness of an item depends directly on the purpose for which a test is being used. For example, a science item that is differentially difficult for women may be judged to be fair in a test designed for certification of science teachers because the item measures a topic that every entry-level science teacher should know. However, that same item, with the same DIF value, may be judged to be unfair in a test of general knowledge designed for all entry-level teachers. (p. 340)

9.3.3 Scaling

Scales based on item response theory (IRT) were derived for each subject area. A single scale was used for summarizing long-term trends at each grade level in each of the subject areas. Three scales were created for reading main assessment data; four scales were created for U.S. history assessment data; and three scales were created for geography assessment data. NAEP uses the methodology of multiple imputations (plausible values) to estimate characteristics of the proficiency distributions. Chapter 11 describes in detail the theoretical underpinnings of NAEP's scaling methods and the required estimation procedures. The basic analysis steps are outlined here.

1. Use the NAEP-BILOG/PARSCALE computer program (described in Chapter 11) to estimate the parameters of the item response functions on an arbitrary scale. This program uses an IRT model incorporating the two- and three-parameter logistic forms used in previous assessments for dichotomously scored items and the generalized partial credit form for polytomously scored items. In order to select starting values for the iterative parameter-estimation procedure for each dataset, the program is first run to convergence, imposing the condition of a fixed normal prior distribution of the proficiency variable. Once these starting values are computed, the main estimation runs model ability as a multinomial distribution That is, no prior assumption about

the shape of the proficiency distribution is made. In analyses involving more than one population, estimates of parameters are made with the overall mean and standard deviation of all subjects' proficiencies specified to be 0 and 1, respectively.

2. Use a version of the MGROUP program (described in Chapter 11), which implements the method of Mislevy (see Chapter 11 or Mislevy, 1991) to estimate proficiency distributions for each student on an arbitrary scale, based on the item parameter estimates and the student's responses to cognitive items and background questions.

3. Determine the appropriate metric for reporting the results and transform the results as needed. This includes the definition of composite scores for reading, U.S. history, and geography, and the linking of current scales to scales from the past, as well as the selection of the mean and variance of new scales.

4. Use random draws from these proficiency distributions (plausible values, in NAEP terminology) for computing the statistics of interest, such as means for demographic groups.

5. Use the jackknife procedure to estimate the standard errors of the mean proficiencies for the various demographic groups.

As explained in Chapter 11, the plausible values obtained through the IRT approach are not optimal estimates of individual proficiency; instead, they serve as intermediate values to be used in estimating subpopulation characteristics. Under the assumptions of the scaling models, these subpopulation estimates will be statistically consistent, which would not be true of subpopulation estimates obtained by aggregating optimal estimates of individual proficiency.

➪ Scaling the Cognitive Items

The data from both the trend and the main assessment samples were scaled using IRT models. For dichotomously scored items two- and three-parameter logistic forms of the model were used, while for polytomously scored items the generalized partial credit model form was used. These two types of items and models were combined in the NAEP scales. The long-term trend and the main assessment data were scaled separately. As described above, in the course of the scaling, analyses of model fit to detect and correct aberrations related to differential functioning across subpopulations, or item-by-time interactions, were carried out. Item parameter estimates on a provisional scale were obtained using the NAEP BILOG/PARSCALE program. The fit of the IRT model to the observed data was examined within each scale by comparing the empirical item response functions with the theoretical curves, as described in Chapter 11.

Long-Term Trend Scaling. Item parameters for reading, mathematics, science, and writing trends were reestimated, separately for each age group using the data from the 1992 assessment as well as data from the 1994 assessment. The resulting scales, based on these reestimated item parameters, were then linked to the existing long-term trend scales.

Scaling with IRT models had not been used for writing in assessments prior to 1992 because software incorporating models for polytomously and dichotomously scored items was not yet available. Hence, in 1992 item parameters for this subject were estimated using the data from the 1984, 1988, and 1990 assessments as well as that from the 1992 assessment. In 1994, data from the 1992 and 1994 assessments were scaled together.

Main Assessment Scaling. Each of the main assessments of reading, U.S. history, and geography have special characteristics that determine the procedures that were followed for main assessment scaling and that determine the relationship between the resultant scales and previously created scales for the subject area. For reading, the key consideration was the degree of similarity between the 1994 assessment of the subject area and earlier assessments in terms of the populations assessed and the characteristics of the assessment instrument used.

The frameworks for the different subject areas dictate differences in the numbers of scales. For reading, U.S. history, and geography, item parameter estimation was performed separately for each of several scales defined in their frameworks, using data from each age/grade sample separately.

A complexity introduced by the 50-minute blocks in each of the subject areas is that those blocks of items must be linked in some way to the shorter blocks. This is complicated by the fact that no students received the shorter blocks in addition to the 50-minute blocks. Because the samples of students receiving each booklet are representative of the population as a whole, it was assumed that the distribution of student proficiency was the same for the students receiving the 50-minute blocks as for the students receiving the booklets containing the shorter blocks.

➪ Generation of Plausible Values for Each Scale

After the scales were developed, plausible values were drawn from the predictive distribution of proficiency values for each student (this process is called *conditioning*). For the long-term trend scales, the plausible values were computed separately for each age or age/grade group and year, and were based on the student's responses to the items going into the scale as well as on the values of a set of background variables that were important for the reporting of proficiency scores. For the reading, U.S. history, and geography main scales, vectors of multivariate plausible values were drawn from the joint distribution of proficiency values for the assessed student. These multivariate plausible values were computed separately for each age/grade and reflected the dependency between scale proficiencies by utilizing shared variation among the scales. All plausible values were later rescaled to the final scale metric using appropriate linear transformations.

The variables used to calculate plausible values for a given main assessment scale or group of scales included a broad spectrum of background, attitude, and experiential variables and composites of such variables and explicitly included all standard reporting variables. The set of variables used in this way were defined with the aim of holding to low levels secondary biases in analyses involving a broad range of variables not included in the conditioning model. To minimize potential convergence problems for the main assessment scales, the original background variables were standardized and transformed into a set of linearly independent variables by extracting principal components from the correlation matrix of the original contrast variables. The principal components, rather than the original variables, were used as independent variables to calculate plausible values for those scales. Trend scales used the same or similar sets of conditioning variables that were used when the scales were originally constructed. Details of the conditioning process and of the NAEP-BGROUP and NAEP-CGROUP computer programs that implement the process are presented in Chapter 11.

➪ Transformation to the Reporting Metric

Reading short-term trend and reading, mathematics, science, and writing long-term trend scales were linked to previous assessment scales via common population linking procedures described in the subject-specific data analysis chapters. Essentially, the 1992 and 1994 data were calibrated together. Data from the two assessments were scaled together in the same BILOG/PARSCALE run, specifying the samples for each assessment as coming from different populations. For each scale, the mean and standard deviation of the 1992 data from this joint calibration were matched to the mean and standard deviation of the 1992 data as previously reported. This then linked the 1994 data to the previously established scale. New scales were established for the U.S. history and geography main assessments. The scales within each age/grade were linked to the other age/grade levels using a revision of the Stocking-Lord procedure for placing item parameters from different IRT scaling runs onto the same scale. Then the metrics for the newly established scales were set to have a mean of 250 and a standard deviation of 50.

The transformations were of the form

$$\theta_{proficiency} = A \bullet \theta_{calibrated} + B$$

where

$\theta_{proficiency}$ = scale level in terms of the system of units of the final scale used for reporting;

$\theta_{calibrated}$ = scale level in terms of the system of units of the provisional NAEP-BILOG/PARSCALE scale;

A = $SD_{new} / SD_{calibrated}$;

B = $M_{new} - A \bullet M_{calibrated}$;

SD_{new} = the estimated or selected standard deviation of the proficiency distribution to be matched;

$SD_{calibrated}$ = the estimated standard deviation of the sample proficiency distribution on the provisional NAEP-BILOG/PARSCALE scale;

M_{new} = the estimated or selected mean of the proficiency distribution to be matched; and

$M_{calibrated}$ = the estimated mean of the sample proficiency distribution on the provisional NAEP-BILOG/PARSCALE scale.

After the plausible values were linearly transformed to the new scale, any plausible value less than 1 was censored to 1 and any value greater than 500 was censored to 500. Fewer than 1 percent of the students in any sample were censored in this way. The final transformation coefficients for transforming each provisional scale to the final reporting scale are given in subsequent chapters.

➪ Definition of Composites for the Multivariate Scales

In addition to the plausible values for each scale, composites of the individual reading, U.S. history, and geography main assessment scales were created as measures of overall proficiency within these subject areas. These composites were weighted averages of the plausible values of the individual scales. The weights reflected the relative importance of the scales and were provided in the frameworks developed by the subject area committees. The weights are approximately proportional to the number of items in each scale at a given age/grade level.

➪ Tables of Proficiency Means and Other Reported Statistics

Proficiencies and trends in proficiencies were reported by age or grade for a variety of reporting categories. Additionally, for the main assessments, the percentages of the students within each of the reporting groups who were at or above achievement levels were reported to provide information about the distribution of achievement within each subject area. All estimates based on proficiency values have reported variances or standard errors and the variances of all estimates based on proficiency values included the error component due to the latency of proficiency values of individual students as well as the error component due to sampling variability.

9.3.4 Drawing Inferences from the Results

Before discussing an observed difference in means, it is appropriate to test whether it is large enough to be statistically significant, or could reasonably have arisen from sampling error alone. If the distributions are normal and the observed difference divided by a weighted function of the two group standard errors exceeds the two-sided critical value 1.96, then a difference that large would arise by chance in only $\alpha = .05$ of cases with these standard errors, in which the true group means were equal.

However, if we wish to make a number of comparisons in the same analysis, say White students versus Black, Hispanic, Asian/Pacific Island, and American Indian students, the probability of finding "significance" by chance for at least one comparison increases with the number (family size) of comparisons. By the Bonferroni inequality, for a family size of 4, for example, the probability of a false positive (Type I error) using $\alpha = .05$ is less than or equal to $4 \times .05 = .20$, larger than most decision-makers would accept.

One general method for controlling error rates in multiple comparisons is to apply this inequality and divide α by the family size. Now $\alpha' = .05/4 = .0125$, and using α', the combined probability of one or more errors in the four comparisons remains controlled at less than or equal to .05. Note that dividing the probability by n is not the same as multiplying the critical value or the confidence band by n. Indeed, in moving from a family size of 1 to 4, we increase the critical value only from 1.960 to 2.498, a 27.4 percent increase. Doubling the family size again, to 8, increases the critical value to 2.735, an additional 9.5 percent increase. To double the initial critical value to 3.92, the family size would have to be increased to 564.

The power of the tests thus depends on the number of comparisons planned. There may be cases where, before the data is seen, it is determined that only certain comparisons will be conducted. As an example, with the five groups above, interest might lie only in comparing the first group with each of the others (family size 4), rather than comparing all possible pairs of groups (family size 10). This means that some possibly significant differences will not be found or discussed, but the planned comparisons will have greater power to identify real differences when they occur.

For trends extending over several administrations, power is gained by testing least-squares fitted linear and quadratic trends, rather than individual pairs of data points. For example, if the linear regression coefficient is significantly greater than 0, and the quadratic coefficient is not different from 0, the trend over time is positive, even though the Bonferroni test might declare no individual pair of points significantly different.

The linear and quadratic components of the trend in average proficiency for a given subject area and age group were estimated by applying two sets of contrasts to the set of average proficiencies by year. The linear component of the trend was estimated by the sum $b_1 = \Sigma\, c_j\, X_j$, where the X_j are the proficiency means by year and the c_j are defined so that b_1 corresponds to the slope of an unweighted regression of the proficiency means on the assessment year. The quadratic component was estimated by the sum $b_2 = \Sigma\, d_j\, X_j$, where the d_j are orthogonal to the c_j and are defined such that b_2 is the quadratic term in the unweighted regression of the proficiency means on the assessment year and the square of the assessment year. The statistical significance of b_1 and b_2 was evaluated by comparing each estimate to its standard error. The standard error of b_1 was computed as the square root of the sum $\Sigma\, c_j^2\, SE_j^2$, where SE_j is the standard error of X_j. The standard error of b_2 was analogously defined. Tests for linear and quadratic trends allow conclusions about the patterns in the means for several points in time or for several related subgroups of students.

9.3.5 Minimum Sample Sizes for Reporting Subgroup Results

In NAEP reports and data summaries, estimates of quantities such as composite and content area proficiency means, percentages of students at or above the achievement levels, and percentages of students indicating particular levels of background variables (as measured in the student, teacher, and school questionnaires) are reported for the total population as well as for key subgroups determined by the background variables. In some cases, sample sizes were not large enough to permit accurate estimation of proficiency and/or background variable results for one or more of the categories of these variables.

For results to be reported for any subgroup, a minimum sample size of 62 was required. This number was arrived at by determining the sample size required to detect an effect size of 0.5 with a probability of .8 or greater.[2] The effect size of 0.5 pertains to the "true" difference in mean proficiency between the subgroup in question and the total population, divided by the standard deviation of proficiency in the total population. In addition, subgroup members must represent at least five primary sampling units (PSUs).

---

[2] A design effect of 2 was assumed for this purpose, implying a sample design-based variance twice that of simple random sampling. This is consistent with previous NAEP experience (Johnson & Rust, 1992).

9.3.6 Estimates of Standard Errors with Large Mean Squared Errors

Standard errors of mean proficiencies, proportions, and percentiles play an important role in interpreting subgroup results and comparing the performances of two or more subgroups. The jackknife standard errors reported by NAEP are statistics whose quality depends on certain features of the sample from which the estimate is obtained. In certain cases, typically when the number of students upon which the standard error is based is small or when this group of students all come from a small number of participating schools, the mean squared error associated with the estimated standard errors may be quite large. In the summary reports, estimated standard errors subject to large mean squared errors are followed by the symbol "!".

The magnitude of the mean squared error associated with an estimated standard error for the mean or proportion of a group depends on the coefficient of variation (*CV*) of the estimated size of the population group, denoted as *N*. This coefficient of variation is estimated by:

$$CV(\hat{N}) = \frac{SE(\hat{N})}{\hat{N}}$$

where $\hat{N}$ is a point estimate of N and $SE(\hat{N})$ is the jackknife standard error of $\hat{N}$.

Experience with previous NAEP assessments suggests that when this coefficient exceeds 0.2, the mean squared error of the estimated standard errors of means and proportions based on samples of this size may be quite large. Therefore, the standard errors of means and proportions for all subgroups for which the coefficient of variation of the population size exceeds 0.2 are followed by "!" in the tables of all summary reports. These standard errors, and any confidence intervals or significance tests involving these standard errors, should be interpreted with caution. (Further discussion of this issue can be found in Johnson & Rust, 1992.)

9.3.7 Teacher Questionnaires

Teachers of students who were in the 4th- and 8th-grade reading, U.S. history, and geography main assessment samples were asked to complete a two-part questionnaire. The first part of the questionnaire pertained to the teacher's background and training (Parts I and IIA in Chapter 2). The second part pertained to the procedures used by the teacher for *specific classes* containing assessed students (Parts IIB I Chapter 2). See Chapter 2 for a description of the teacher questionnaires. It is important to note that because geography is not often taught in a class separate from U.S. history or the social sciences, teachers were often confused by the request for information about classes containing students in geography.

To analyze the data from the teacher questionnaires with respect to the students' data, each teacher's questionnaire had to be matched to all of the sampled students who were taught by that teacher.

In the subsequent chapters two separate match rates for each grade are given. The first is the percentage of students that could be matched to both the first and second parts of the teacher questionnaire. For these students, information is available not only about the background and training of their teachers, but also about the methods used in the particular class they attended. The second match rate is the percentage of students that could be matched to the first part of the teacher questionnaire. This match rate is larger because more students could be matched with information about a teacher than with information about the particular class they attended. Note that these match rates only reflect the student-level missing data. They do not reflect the additional missing data due to item-level nonresponse on the part of teachers. Variables derived from the teacher questionnaires were used as reporting variables at the student level and as variables that contributed to conditioning for the appropriate samples.

Teachers of students who were in grade 4 reading, U.S. history, and geography main assessment samples were asked to complete a two-part questionnaire. As with the grade 8 teacher questionnaire, the first part pertained to the teacher's background and training. Unlike the grade 8 teacher questionnaire, the second part pertained to only a single class that the teacher taught. In development of the questionnaires, it was thought that fourth-grade teachers would teach one class in each subject. In practice, that was found to be untrue for a number of teachers. A single student-teacher match rate matching students to the first part of the questionnaire is reported for grade 4 in the following chapters.

9.3.8 Dimensionality Analysis

Over the years a number of studies have been conducted in order to seek answers to the question of how many dimensions underlie the various NAEP assessment instruments, and whether there is a sufficiently strong first dimension to support inferences about a composite scale in those subject matter areas in which the cognitive assessment instruments comprise more than one scale (mathematics and reading). These studies are primarily meant to support the dimensions defined by the framework for the subject area studied. Findings from those studies (briefly reviewed below) can, for the most part, be applied to current NAEP instruments because there is a heavy emphasis on keeping the frameworks consistent across years. From time to time, however, changes do take place and in cases where this occurs it is necessary to conduct studies to determine the effects of such changes on the dimensionality of the item pools. For the 1992 assessment such a study was conducted (Carlson, 1993) to determine whether the increasing emphasis on extended constructed-response items that are scored polytomously has any effect on the dimensionality. It was determined that for the 1992 NAEP data, item type was not related to any of the dimensions identified.

➪ Previous Dimensionality Analyses of NAEP Data

NAEP reading assessment data collected during the 1983-84 academic year was studied for dimensionality by Zwick (1986, 1987) who also examined simulated data designed to mirror the NAEP reading item-response data but having known dimensionality. Principal components analysis was applied to both phi and tetrachoric correlation matrices, and full information item factor analysis (Bock & Aitkin, 1981; Bock, Gibbons, & Muraki, 1988) implemented in the TESTFACT computer program (Wilson, Wood, & Gibbons, 1991) was applied to portions of the dataset, as were Rosenbaum's (1984, 1985) dimensionality testing procedures. Analysis of the simulated datasets allowed her to determine whether the BIB spiraling design artificially increases dimensionality. Zwick found substantial agreement among the various statistical procedures, and that the results using BIB spiraling were similar to results for complete datasets. Overall she concluded that "it is not unreasonable to treat the data as unidimensional (1987, p. 306)."

The topic of Rock's (1991) investigation was "whether the presently reported scale scores do span a multidimensional space defined by the content area scales at each of the three grade levels in mathematics and science (p. 1)." He formed two parcels of items that are homogeneous with respect to content for each subtest of the NAEP mathematics and science tests from the 1990 assessment, and studied their dimensionality using confirmatory factor analysis. The resulting factor intercorrelations averaged across booklets ranged from .86 to .95 in mathematics, and from .94 to .96 in science. Rock's conclusion was that there was little evidence for discriminant validity except for the geometry scale at the eighth-grade level, and that "we are doing little damage in using a composite score in mathematics and science (p. 2)."

A second-order factor model was used by Muthén (1991) in a further analysis of Rock's mathematics data, to examine subgroup differences in dimensionality. Evidence of content-specific variation within subgroups was found, but the average (across seven booklets) percentages of such variation was very small, ranging from essentially 0 to 22, and two-thirds of these percentages were smaller than 10.

Carlson and Jirele (1992) used the same full information item factor analysis procedure used by Zwick, and normal harmonic factor analysis (McDonald, 1962, 1967, 1981) as implemented in the NOHARM program (Fraser, 1988) to examine 1990 NAEP mathematics data at three grade levels. Analyses of simulated one-dimensional data were also conducted, and the fit to these data, as measured by the root mean square residual and the Akaike Information Coefficient (Akaike, 1987) was slightly better than that to the real NAEP data. The simulated data were generated using a three-parameter IRT model and a BIB spiraling design like that used in NAEP. Although there was some evidence suggesting more than one dimension in the NAEP data, the strength of the first dimension led the authors to conclude that the data "are sufficiently unidimensional to support the use of a composite scale for describing the NAEP mathematics data, but that there is evidence that two dimensions would better fit the data than one (p. 31)."

➪ Analyses of the 1992 NAEP Cognitive Instruments

As mentioned above, the dimensionality analyses conducted on the 1992 assessment data were directed towards study of the question as to whether dimensionality is affected by inclusion of polytomously scored items on the 1992 NAEP cognitive instruments. Results from the 1992 analyses for reading are related to the 1994 reading assessment, since about two-thirds of the 1994 NAEP cognitive instruments are the same as those used in the 1992 assessment of reading. It should be noted that several of the procedures used by previous researchers could not be used for the 1992 or 1994 data because they do not allow for polytomously scored items.

Carlson (1993) used the LISREL computer program, employing a maximum likelihood parameter estimation technique to study the factor structure of 1994 mathematics and reading cognitive instruments. Results were evaluated, as suggested by McDonald (1981, 1982a, 1982b, 1985), through use of residuals from the fitted model. The square root of the mean squared residual was the statistic used.

It should be noted that in the 1992 NAEP instruments used in this study there is only one polytomously scored item in each block of the BIB spiral. Hence there were only three polytomously scored items in each student's responses. Additionally, at the twelfth grade there was one block of mathematics items that had no polytomously scored item so the students assigned booklets containing that block were administered only two such items. In order to revisit one of the questions studied by Zwick (1986, 1987) complete datasets were simulated as well as datasets using the BIB design.

Mathematics Results. As revealed in the values of the root mean square residual statistics there was no obvious difference in the fit with one, two, or three factors at the twelfth-grade level. At the lower grade levels there was some decrease in the root mean square residual when more factors were fitted, but the decrease was so minimal that the data were concluded to be essentially unidimensional. Types of items, one of the primary focuses of the research, do not appear to result in multidimensionality in the context of the types of structures in the NAEP mathematics data. That is, there were only minor differences between one-dimensional solutions and a two-dimensional solution in which the second dimension is defined by the polytomously scored (or constructed-response) items, and the first by the dichotomously scored (or multiple-choice) items.

Reading Results. Similar analyses were conducted on the 1992 NAEP reading assessment data. The results summarized above are similar to those for the 1992 NAEP reading assessment. In the case of reading the lower correlations in the actual data suggested studying more than one simulated factor structure. Because of the specific blocks assembled into the NAEP reading instruments, the actual data used in this study never included items measuring more than one of the three NAEP reading scales. Each block, however, as was pointed out above, consists of a reading passage and several items (9 to 13) about that passage. Hence the multidimensional simulated data were generated as if each passage defined a separate dimension. The correlations among the actual reading scales that were used in generating these multidimensional data were lower than those among the mathematics scales.

In the actual data, fitting more than one factor had more effect on the size of the root mean square residual statistics and interfactor correlations than was the case in mathematics, at least at the eighth- and twelfth-grade levels. Again, however, there appeared to be little or no effect associated with item type—dichotomously scored versus polytomously scored, or multiple-choice versus constructed-response. In the case of simulating a complete data matrix of three dimensions at the twelfth-grade level, the root mean square residual statistic did seem to indicate some lack of fit when one or two dimensions were fitted rather than the three that underlie the generation process. The trend in the actual twelfth-grade data showed less of an effect than in the simulated data, suggesting fewer than three dimensions in the NAEP reading instruments.

Conclusions. The relative sizes of the root mean square residual statistics for the simulated as compared to actual data suggested that lack of fit may be more due to the BIB spiraling design of NAEP than the number of dimensions fitted. Kaplan (1995) similarly found that the chi-squared goodness of fit statistic in the maximum likelihood factor analysis model was inflated when data were generated using a BIB design. As might be expected, it was found that the sizes of the root mean square residual statistics for the incomplete simulation condition (a BIB design as in the actual NAEP assessment) were more like those of the real data than were those of the case of simulation of a complete data matrix. Consistent with findings of Zwick (1986, 1987), however, the incomplete design for data collection used in NAEP does not appear to be artificially inflating the number of dimensions identified using these procedures.

9.4 OVERVIEW OF CHAPTERS 10 THROUGH 19

The remaining chapters in Part II of this report are as follows:

***Chapter 10*:** The 1994 National Assessment used a stratified multistage probability sampling design that provided for sampling certain subpopulations at higher rates (see Chapter 3). Because probabilities of selection are not the same for all assessed students, sampling weights must be used in the analysis of NAEP data. Also, in NAEP's complex sample, observations are not independent. As a result, conventional formulas for estimating the sampling variance of statistics are inappropriate. Chapter 10 describes the weighting procedures and methods for estimating sampling variance that are necessitated by NAEP's sample design. Further detail on sampling and weighting procedures is provided in *1994 National Assessment of Educational Progress Sampling and Weighting Procedures, National Assessment, Final Report* (Wallace & Rust, 1996), a report prepared by Westat, Inc., the NAEP subcontractor in charge of sampling.

***Chapter 11*:** A major NAEP innovation introduced by ETS is the reporting of subject-area results in terms of IRT-based scales. Scaling methods can be used to summarize results even when students answer different subsets of items. For purposes of summarizing item responses, NAEP developed a scaling technique that has its roots in IRT and in the theories of imputation of missing data. Chapter 11 describes this scaling technique, the underlying theory, and the application of these methods to 1994 NAEP data. The final section of Chapter 11 gives an overview of the NAEP scales that were developed for the 1994 assessment.

***Chapter 12*:** The main short-term trend component of the 1994 reading analysis is described in this chapter. A detailed analysis of the main assessment of reading was conducted for grades 4, 8, and 12, including a study of the association between reading proficiency and student background variables. The results from this component of the analysis can be compared with results from the 1992 reading analysis to examine short-term trends. At grades 4 and 8, background information and data on instructional methods were collected from teachers and the relation of these variables to reading proficiency was examined. The main assessment analyses are reported in *1994 NAEP Reading: A First Look—Findings from the National Assessment of Educational Progress* (Williams, Reese, Campbell, Mazzeo, & Phillips, 1995) and *NAEP 1994 Reading Report Card for the Nation and the States: Findings from the National Assessment of Educational Progress and Trial State Assessment* (Campbell, Donahue, Reese, & Phillips, 1996). A special study of choice of reading passage, the *NAEP Reader*, was also conducted.

***Chapter 13*:** The main assessment analysis of the U.S. history data is detailed in Chapter 13. This analysis included a study of the association of U.S. history knowledge with instructional techniques and student background variables. At grades 4 and 8, background information and data on instructional methods were collected from teachers and the relation of these variables to U.S. history proficiency was examined. The U.S. history results appear in *NAEP 1994 U.S. History: A First Look, Findings from the National Assessment of Educational Progress* (Williams, Lazer, Reese, & Carr, 1995) and the *NAEP 1994 U.S. History Report Card: Findings from the National Assessment of Educational Progress* (Beatty, Reese, Persky, & Carr, 1996). There was also a pilot study of students' U.S. history skills evaluated as they participated in groups with specific history-related tasks.

Chapter 14: The main assessment analysis of the geography data is detailed in Chapter 14. This analysis included a study of the association of geography knowledge with instructional techniques and student background variables. At grades 4 and 8, background information and data on instructional methods were collected from teachers and the relation of these variables to geography proficiency was examined. The geography results appear in *NAEP 1994 Geography: A First Look—Findings from the National Assessment of Educational Progress* (Williams, Reese, Lazer, & Shakrani, 1995) and the *NAEP 1994 Geography Report Card: Findings from the National Assessment of Educational Progress* (Persky, Reese, O'Sullivan, Lazer, Moore, & Shakrani, 1996).

Chapter 15: The reading trend results for the years 1971 through 1992 were extended to include 1994 at ages 9, 13, and 17. The results of the reading trend analysis, which include the percentages of students at or above the reading scale anchor points established in 1984, are reported in *NAEP 1994 Trends in Academic Progress: Achievement of U.S. Students in Science, 1969 to 1994; Mathematics, 1973 to 1994; Reading, 1971 to 1994; and Writing, 1984 to 1994* (Campbell, Reese, O'Sullivan, Dossey, & Donahue, 1996).

Chapter 16: The long-term trend assessment analysis of the mathematics data is detailed in Chapter 16. The results of the trend analysis, which provided links from 1976 through 1994 for ages 9, 13, and 17, are reported in *NAEP 1994 Trends in Academic Progress.*

Chapter 17: The long-term trend assessment analysis of the science data is described in Chapter 17. The science trend results, which provide a link to 1970, 1973, 1977, 1982, 1986, 1990, and 1992, are reported in *NAEP 1994 Trends in Academic Progress.*

Chapter 18: Like the 1994 mathematics and science assessments, the writing assessment in 1994 consisted of only one component, a long-term trend. The writing trend results, which provide a link to 1984, 1988, 1990, and 1992 for grades 4, 8, and 12, are reported in *NAEP 1994 Trends in Academic Progress.*

Chapter 19: This chapter presents basic data from the 1994 assessment, including the properties of the measuring instruments and characteristics of the sample.

Chapter 10

WEIGHTING PROCEDURES AND ESTIMATION OF SAMPLING VARIANCE[1]

Eugene G. Johnson
Educational Testing Service

Leslie Wallace, Keith F. Rust, and Ellen Tubbs
Westat, Inc.

10.1 INTRODUCTION

As was the case in previous assessments, the 1994 National Assessment used a complex sample design with the goal of securing a sample from which estimates of population and subpopulation characteristics could be obtained with reasonably high precision (as measured by low sampling variability). At the same time, it was necessary that the sample be economically and operationally feasible to obtain. The resulting sample had certain properties that had to be taken into account to ensure valid analyses of the data from the assessment.

The 1994 NAEP sample was obtained through a stratified multistage probability sampling design that included provisions for sampling certain subpopulations at higher rates (see Chapter 3). To account for the differential probabilities of selection, and to allow for adjustments for nonresponse, each student was assigned a sampling weight. Section 10.2 discusses the procedures used to derive these sampling weights.

Another consequence of the NAEP sample design is its effect on the estimation of sampling variability. Because of the effects of cluster selection (students within schools, schools within primary sampling units) and because of the effects of certain adjustments to the sampling weights (nonresponse adjustment and poststratification), observations made on different students cannot be assumed to be independent of one another. As a result, ordinary formulas for the estimation of the variance of sample statistics, based on assumptions of independence, will tend to underestimate the true sampling variability. Section 10.3 discusses the jackknife technique used by NAEP to estimate sampling variability. (The estimation of variability due to imperfect measurement of individual proficiency is discussed in Chapter 11.)

The jackknife technique provides good quality estimates of sampling variability but requires considerable computations. Section 10.4 suggests the use of design effects, combined with conventional variance estimation formulas, as a simple approximation to sampling variability.

---

[1] The statistical programming for this chapter was provided by Phillip Leung of Educational Testing Service and Brice Hart, Marianne Whitlock, Ngoan Vo, Fran Cohen, and John Edmonds of Westat, Inc.

Since the sample design determines the derivation of the sampling weights and the estimation of sampling variability, it will be helpful to note the key features of the 1994 NAEP sample design. A description of the design appears in Chapter 3, and the various assessment instruments are detailed in Chapter 4.

The 1994 sample was a multistage probability sample consisting of four stages of selection. The first stage of selection, the primary sampling units (PSUs), consisted of counties or groups of counties. The second stage of selection consisted of elementary and secondary schools. The assignment of sessions to sampled schools comprised the third stage of sampling, and the fourth stage involved the selection of students within schools and their assignment to sessions. The probabilities of selection of the first-stage sampling units were proportional to measures of their size, while the probabilities for subsequent stages of selection were such that the overall probabilities of selection of students were approximately uniform, with exceptions for certain subpopulations that were oversampled by design. For the main assessment, schools with relatively high concentrations of Black students and/or Hispanic students were deliberately sampled at twice the normal rate to obtain larger samples of respondents from those subpopulations, in order to increase the precision in the estimation of the characteristics of these subpopulations. Also for the main assessment, nonpublic school students were sampled at three times the normal rate, again so as to increase the precision of estimates for this population subgroup. For all assessment components, students from schools with smaller numbers of eligibles received lower probabilities of selection, as a means of enhancing the cost efficiency of the sample.

The 1994 main assessment includes three student cohorts: students who were *either* in the fourth grade *or* 9 years old; students who were *either* in the eighth grade *or* 13 years old; and students who were *either* in the twelfth grade *or* 17 years old. The main assessment represents two overlapping samples of students, the first of specified grades (of any age) and the second of specified ages (in any grade). Students were age-eligible if they were born in the appropriate calendar year (1984, 1980, or 1976). Each student cohort is called an "age class." The main assessment of all age classes was conducted in the spring of 1994.

The full 1994 assessment also includes a number of additional samples that used the age definitions, times of testing, and modes of administration used in previous assessments. These are referred to as long-term trend samples. The purpose of these samples was to provide the statistical linkage between the 1994 data and data from previous assessments.

The full 1994 NAEP assessment thus includes a number of different samples from several populations. Each of these samples has its own set of weights that are to be used to produce estimates about the characteristics of the population addressed by the sample (the target population). The various samples and their target populations are as follows:

The Main Samples of Students. These samples, one for each of the three age classes, were drawn in January through March of 1994, using the calendar-year age definitions, and consist of all students assessed in the main assessment. The target population for each of these samples consists of all students who were of the specified age or in the specified grade and were deemed assessable by their school. For each age class, there were three distinct assessments—reading, U.S. history, and geography—conducted in two distinct session types. The U.S. history and geography assessments were administered in the same session.

To facilitate separate analyses, a set of weights was produced corresponding to each age class and assessment type. Thus in total there were nine sets of final student weights for the sample of assessed students in the main assessment.

Long-Term Trend Samples for Reading and Writing. These consist of samples comparable to the 1984 main assessment and address the subject areas of reading and writing. The samples were collected by grade and age for age 9/grade 4, age 13/grade 8, and age 17/grade 11, using the age definitions and time of testing from 1984. As in that assessment, print administration was used. Six assessment booklets were administered at each age class. The respondents to the combined set of assigned booklets at a given age class constitute a representative sample of the population of students who were in the specified grade *or* of the specified age. The respondents to any one of the booklets also constitute a representative sample.

Long-Term Trend Samples for Mathematics and Science. These consist of samples comparable to those used for the measurement of trends in 1986. The samples were collected by age only and using the same age definitions and time of testing as in the long-term trend assessment in 1986. As in that assessment, the administration of mathematics and science questions was paced with an audiotape. For ages 9 and 13, three assessment booklets were administered to each age group while two booklets were administered at age 17. The respondents to any one of the booklets assigned to a given age constitute a representative sample of the population of all students of that age.

For purposes of sampling and weighting, the assessment samples are categorized as "tape-administered" or "print-administered" according to whether or not paced audiotapes were used in the administration:

1) *Tape-administered samples* are samples that required audiotape pacing in the assessment (the long-term trend assessments in mathematics and science). For these samples, all students within a particular assessment session receive the same booklet and were paced through at least part of the booklet with an audiotape.

2) *Print-administered samples* are all main assessment samples and the long-term trend assessments of reading and writing. For these samples, no audiotape pacing was employed and the assessment booklets were spiraled through each assessment session (that is, the booklets were systematically interspersed and assigned for testing in that order).

10.2 DERIVATION OF THE SAMPLE WEIGHTS

As indicated previously, NAEP uses differential sampling rates, deliberately oversampling certain subpopulations to obtain larger samples of respondents from those subgroups, thereby enhancing the precision of estimates of characteristics of these oversampled subgroups. As a result of the oversampling, these subpopulations, corresponding to students from schools with high concentrations of Black and/or Hispanic students, and from nonpublic schools, are overrepresented in the sample. Lower sampling rates were introduced also for very small schools (those schools with only 1 to 19 eligible students). This reduced level of sampling from small schools was undertaken in an approximately optimum manner as a

BEST COPY AVAILABLE

means of reducing variances per unit of cost. Appropriate estimation of population characteristics must take disproportionate representation into account. This is accomplished by assigning a weight to each respondent, where the weights approximately account for the sample design and reflect the appropriate proportional representation of the various types of individuals in the population.

The weighting procedures for 1994 included computing the student's base weight, the reciprocal of the probability that the student was selected for a particular session. Such weights are those appropriate for deriving estimates from probability samples via the standard Horvitz-Thompson estimator (see Cochran, 1977, Section 9A.7). These base weights were adjusted for nonresponse and then subjected to a trimming algorithm to reduce a few excessively large weights. The weights were further adjusted by a poststratification procedure in an effort to reduce the sampling error and certain potential biases of estimates relating to student populations corresponding to several subgroups of the total population. Poststratification was performed by adjusting the weights of the sampled students so that the resulting estimates of the total number of students in a number of specified subgroups of the population corresponded to population totals based on information from the Current Population Survey and Census Bureau estimates of the population. The subpopulations were defined in terms of race, ethnicity, geographic region, age, grade, and modal grade status.

The following sections provide an overview of the procedures used to derive the sampling weights. Further details in the derivation of these weights can be found in the *1994 National Assessment of Educational Progress Sampling and Weighting Procedures, National Assessment, Final Report* (Wallace & Rust, 1996).

10.2.1 Student Base Weight

The base weight assigned to a student is the reciprocal of the probability that the student was selected for a particular assessment. That probability is the product of four factors:

1) the probability that the PSU was selected;

2) the conditional probability, given the PSU, that the school was selected;

3) the conditional probability, given the sample of schools in a PSU, that the school was allocated the specified assessment; and

4) the conditional probability, given the school, that the student was selected for the specified assessment.

Thus, the base weight for a student may be expressed as the product

$$W_B = PSUWT \cdot SCHWT \cdot SESSWT \cdot STUSCHW$$

where PSUWT, SCHWT, SESSWT, and STUSCHW are, respectively, the reciprocals of the preceding probabilities.

Variations across the various 1994 assessments in probabilities of selection, and consequently of weights, were introduced by design, either to increase the effectiveness of the sample in achieving its goals of reporting for various subpopulations, or to achieve increased efficiency per unit of cost.

10.2.2 Adjustment of Base Weights for Nonresponse

The base weight for a selected student was adjusted by two nonresponse factors. The first of these was to adjust for sessions that were not conducted, either because a contacted school refused or because a cooperating school did not conduct all of its assigned sessions. The second adjustment was needed to adjust for students who were (or should have been) invited to the assessment but did not appear either in the scheduled session or a makeup session. Thus, the nonresponse adjusted weight for a student is of the form

$$W' = W_B \bullet \text{SESNRF} \bullet \text{STUNRF}$$

where the nonresponse adjustment factors SESNRF and STUNRF are computed as described below. It should be noted that the nonresponse adjustments assume that nonresponse occurs at random within the categories within which adjustments are made (see Little & Rubin, 1987). Some degree of bias could result to the extent that this assumption is false.

10.2.2.1 Session Nonresponse Adjustment (SESNRF)

Sessions were assigned to schools before cooperation status was final. The session nonresponse adjustment was intended to compensate for session nonresponse due to refusing schools or individual sessions not conducted. These factors were computed separately within classes formed by the first three digits of PSU stratum, called subuniverse (formed by crossing the PSU major stratum and the first socioeconomic characteristic used to define the final PSU stratum; see Section 3.2 for more detail). Occasionally, additional collapsing of classes was necessary to improve the stability of the adjustment factors, especially for the smaller assessment components. Most classes needing collapsing contained small numbers of cooperating schools. Occasionally, classes with low response rates were collapsed.

In PSU h, the session nonresponse adjustment factor SESNRF_h was given by

$$SESNRF_{hs} = \frac{\sum_{i \in B'_{hs}} PSUWT_i \cdot SCHWT_i \cdot SESSWT_{hi} \cdot G_i}{\sum_{i \in C_{hs}} PSUWT_i \cdot SCHWT_i \cdot SESSWT_{hi} \cdot G_i}$$

where

PSUWT_i = the PSU weight for the PSU containing school i;

SCHWT_i = the school weight for school i;

SESSWT_{hi} = the session allocation weight for school i for session h;

BEST COPY AVAILABLE

G_i = the estimated number of age- plus grade-eligible students in school i in the case of print-administered sessions, and the estimated number of age-eligible students in the case of the tape-administered sessions, to which only age eligibles were invited (the values of G_i were based on QED data);

set B'_{hs} = consists of all in-scope originally sampled schools allocated to session h in subuniverse s (excluding substitutes); and

set C_{hs} = consists of all schools allocated to session h in subuniverse s that ultimately participated (including substitutes).

10.2.2.2 Student Nonresponse Adjustment (STUNRF)

Student nonresponse adjustment factors were completed separately for each assessment. For assessed students in the long-term trend samples, the student nonresponse adjustment was made separately for classes of students based on subuniverse and modal grade status (at or above modal grade, below modal grade). For assessed students in the main samples, the adjustment classes were based on subuniverse, modal grade status, and race class (White or Asian, other). In some cases, nonresponse classes were collapsed into one to improve the stability of the adjustment factors. For each class *c* in assessment *g* in session *h*, the student nonresponse adjustment factor $STUNRF_{ghc}$ is computed by

$$STUNRF_{ghc} = \frac{\sum_{A'_{ghc}} PSUWT_j \cdot SCHWT_j \cdot SESSWT_{hj} \cdot SESNRF_{hj} \cdot STUSCHW_{ghj}}{\sum_{B'_{ghc}} PSUWT_j \cdot SCHWT_j \cdot SESSWT_{hj} \cdot SESNRF_{hj} \cdot STUSCHW_{ghj}}$$

where

$PSUWT_j$ = the PSU weight for the PSU containing student *j*;

$SCHWT_j$ = the school weight for the school containing student *j*;

$SESSWT_{hj}$ = the session allocation weight for the school containing student *j*;

$SESNRF_{hj}$ = the session nonresponse adjustment factor for the school containing student *j*;

$STUSCHW_{ghj}$ = the within-school student weight for student *j* in session *h* in assessment *g*;

Set A'_{ghc} = consists of the students in class *c* who were sampled for assessment *g* in section *h* and not excluded; and

Set B'_{ghc} = consists of the students in class *c* who were assessed in assessment *g* in session *h*.

10.2.2.3 Evaluation of Potential for Bias Resulting from School and Student Nonparticipation

Although school and student nonresponse adjustments are intended to reduce the potential for nonparticipation to bias the assessment results, they cannot completely eliminate this potential bias with certainty. The extent of bias remains unknown, of course, since there are no assessment data for the nonparticipating schools and students.

Some insight can be gained about the potential for residual nonresponse bias, however, by examining the weighted school- and student-level distributions of characteristics known for both participants and nonparticipants, especially for those characteristics known or thought likely to be related to achievement on the assessment. If the distributions for the full sample of schools (or students) without the use of nonresponse adjustments are close to those for the participants with nonresponse adjustments applied, there is reason to be confident that the bias from nonparticipation is small.

There are several school-level characteristics available for both participating and nonparticipating schools. The tables below show the combined impact of nonresponse and of the nonresponse adjustments on the distributions of schools (weighted by the estimated number of eligible students enrolled) and students, by the type of school (public, Catholic, other nonpublic) the size of the school as measured by the estimated number of eligible students enrolled, and the urban/rural nature of the place where the school is located. Three size classes have been defined for each age class. The data are for the main assessments all session types combined.

Several student-level characteristics are available for both absent and assessed students. The tables that follow show the impact of school nonresponse and nonresponse adjustments, and student nonresponse and nonresponse adjustments on the distributions of eligible students for each age class. The discussion focuses on the reading assessment, since it is the largest. It is assumed that other large assessments would behave similarly. The distributions are presented by age category (two levels), grade category (two levels), race category (White, Black, Hispanic, and other), gender, IEP, and LEP. Age is divided into two categories: at or below modal age, and above modal age. Grade is divided into at or above modal grade, and below modal grade.

Table 10-1 shows the weighted marginal distributions of schools for each of the three classification variables for each age class, using the full sample of in-scope schools—those participating, plus those refusals for which no substitute participated. Table 10-2 shows the same distribution based only on participating schools, with school nonresponse adjustments applied to them. For school-level data, the school nonresponse adjustment is actually a composite of the session nonresponse adjustment factors derived for use with student-level data.

It can be seen from the tables that, even though the level of school nonparticipation is as high as 24 percent for age class 17 (see Table 3-11) and somewhat lower for the other age classes, the distributions for the three characteristics considered remain similar.

Tables 10-3 and 10-4 present similar data for students. Table 10-3 shows the distributions for assessed and absent students (with base weights adjusted for school nonparticipation) while Table 10-4 shows them for assessed students only, with the student nonresponse adjustments also applied to the weights. The rates of student nonparticipation were 6 percent for age class 9, 9 percent for age class 13, and 18 percent for age class 17 (see Table 3-11). The tables show that for the distributions of type of school attended and place where the school is located, the combined effect of student nonparticipation and the subsequent nonresponse adjustments have resulted in very little change in distribution.

Table 10-1

Distribution of Populations of Eligible Students Based on Full Weighted Sample of Eligible Schools, 1994 Main NAEP Samples

Population	Age Class 9	Age Class 13	Age Class 17
Total population	5,093,865	4,698,846	4,215,678
School type			
Catholic	5.0%	5.3%	4.2%
Other Nonpublic	4.6%	4.4%	3.7%
Public	90.3%	90.3%	92.0%
School size*			
1	20.0%	13.5%	5.6%
2	21.1%	37.8%	42.1%
3	58.8%	47.8%	52.2%
School location			
Large city	12.2%	14.7%	12.6%
Midsize city	24.1%	23.5%	20.4%
Urban fringe/large city	25.1%	18.7%	24.7%
Urban fringe/midsize city	14.9%	16.3%	15.0%
Large town	3.0%	2.1%	2.2%
Small town	10.5%	10.4%	11.5%
Rural MSA	3.1%	4.7%	4.7%
Rural nonMSA	7.2%	9.5%	8.9%

**Note*:* *School size = number of eligible students enrolled:*

	1	2	3
Age Class 9	1-49	50-99	100+
Age Class 13	1-49	50-299	300+
Age Class 17	1-49	50-399	400+

Table 10-2

Distribution of Populations of Eligible Students Based on Weighted Sample of Participating Schools with School Nonresponse Adjustments, 1994 Main NAEP Samples

Population	**Age Class 9**	**Age Class 13**	**Age Class 17**
Total population	5,093,865	4,698,846	4,215,678
School type			
Catholic	5.7%	6.7%	4.7%
Other Nonpublic	4.2%	4.1%	3.6%
Public	90.1%	89.2%	91.7%
School size*			
1	20.9%	14.6%	6.5%
2	23.0%	39.4%	40.7%
3	56.1%	46.0%	52.9%
School location			
Large city	12.1%	14.3%	12.4%
Midsize city	24.3%	22.5%	21.3%
Urban fringe/large city	25.4%	18.8%	25.5%
Urban fringe/midsize city	15.5%	17.7%	13.6%
Large town	3.2%	2.1%	2.5%
Small town	10.3%	11.3%	12.3%
Rural MSA	1.8%	4.4%	4.7%
Rural nonMSA	7.3%	9.0%	7.8%

Note: School size = number of eligible students enrolled:

	1	2	3
Age Class 9	1-49	50-99	100+
Age Class 13	1-49	50-299	300+
Age Class 17	1-49	50-399	400+

BEST COPY AVAILABLE

Table 10-3

Distribution of Populations of Eligible Students Based on Assessed and Absent Students from Participating Schools, 1994 Main NAEP Samples

Population	Age Class 9	Age Class 13	Age Class 17
Total population	4,514,692	4,518,079	3,328,622
School type			
Catholic	6.2%	6.1%	4.8%
Other Nonpublic	3.7%	3.8%	3.7%
Public	90.1%	90.1%	91.5%
School location			
Large city	10.5%	12.4%	11.0%
Midsize city	23.5%	22.1%	21.7%
Urban fringe/large city	25.8%	19.1%	25.0%
Urban fringe/midsize city	16.3%	17.0%	14.1%
Large town	3.1%	2.1%	2.1%
Small town	11.0%	11.7%	12.7%
Rural MSA	1.8%	5.3%	4.6%
Rural nonMSA	8.0%	10.4%	8.8%

Table 10-4

Distribution of Populations of Eligible Students Based on Assessed Students from Participating Schools with Student Nonresponse Adjustments Applied, 1994 Main NAEP Samples

Population	Age Class 9	Age Class 13	Age Class 17
Total population*	4,514,687	4,518,078	3,328,632
School type			
Catholic	6.3%	6.4%	5.5%
Other Nonpublic	3.8%	3.9%	4.0%
Public	89.9%	89.7%	90.4%
School location			
Large city	10.2%	12.2%	10.3%
Midsize city	23.7%	21.8%	21.3%
Urban fringe/large city	25.9%	19.3%	25.4%
Urban fringe/midsize city	16.3%	17.2%	14.5%
Large town	3.1%	2.0%	2.0%
Small town	11.0%	11.8%	12.6%
Rural MSA	1.8%	5.3%	4.8%
Rural nonMSA	8.0%	10.3%	8.9%

***Note**: *Totals do not match those in Table 10-3 due to rounding.*

Tables 10-5, 10-6, and 10-7 show that both school and student-level nonresponse and nonresponse adjustments have little effect on the distributions of eligible students. All of the distributions in the tables are similar. Table 10-5 shows the weighted distributions of eligible students in participating schools. The distributions are calculated using the base weights of invited students unadjusted for school-level nonresponse. Table 10-6 shows the same distributions using weights adjusted for school nonresponse. Table 10-7 shows the weighted distributions of eligible students using the student nonresponse adjusted weights of assessed students.

When comparing Tables 10-6 and 10-7, we expect the distributions by age category, grade category, and race/ethnicity to be similar because these variables were used to determine student nonresponse adjustment classes. However, the distributions by gender, IEP, and LEP are also similar. To the extent that nonrespondents would perform like respondents with the same characteristics (defined by the classification variables in the tables), the bias in the assessment data is small.

Further information about potential nonresponse bias can be gained by studying the absent students. NAEP proficiency estimates are biased to the extent that assessed and absent students within the same weighting class differ in their distribution of proficiency. It seems likely that the assumption that absent students are similar in proficiency to assessed students is reasonable for some absent students—namely, those whose absence can be characterized as random. Conversely, it seems likely that students with longer and more consistent patterns of absenteeism—such as truants, dropouts, near dropouts, and the chronically ill—are unlikely to be as proficient as their assessed counterparts.

In the 1994 assessments, schools were asked to classify each absent student into one of nine categories. The results of this classification for the reading assessment are shown in Table 10-8. The discussion focuses on the reading assessment because it is the largest. It is assumed that the other large assessments would behave similarly.

Table 10-8 shows that, as anticipated, the majority of absence from the assessment was the result of an absence from school of a temporary and unscheduled nature. Past NAEP experience has shown that absence among 17 year olds occurs at about three times the rate of absence among 9 year olds. This is also evident in the table. The proportion of absence classified as temporary differs somewhat by age class, but is of the same magnitude for age classes 9 and 17. This suggests strongly that a substantial proportion of the temporary absences among age class 17 students is not a result of illness, because such absences are occurring at almost three times the rate that they do among age class 9 students. Whereas it might be reasonable to regard temporary absence due to illness as independent of proficiency, for other temporary absences, this appears less tenable. The data in the table give support to the contention that, at age class 9, student absences are unlikely to introduce any significant bias into NAEP estimates. The absentee rate is low; most absences are temporary, and one third of the remaining absences are a result of parental refusal.

At age class 17, however, a significant component of absenteeism is contributed by students who are in school on the day of the assessment but failed to attend the session (9.8 percent). Chronic truants, those suspended, and those in school but not invited, constitute the obvious candidates for potential bias. As they comprise only 4.1 percent of absent students, their potential for introducing significant bias under the current procedures is minor.

Table 10-5

Distribution of Populations of Eligible Students in Participating Schools Based on Invited Students from Participating Schools Using Base Weights Unadjusted for School Nonresponse
1994 Main NAEP Reading Sample

Population	Age Class 9	Age Class 13	Age Class 17
Total population	3,913,006	3,771,786	2,564,845
Age category			
At modal age or younger	73.0%	71.0%	75.3%
Older than modal age	27.0%	29.0%	24.7%
Grade category			
At or above modal grade	75.6%	73.4%	71.2%
Below modal grade	24.4%	26.6%	28.8%
Age category, grade category			
At or younger, at or above	48.6%	44.4%	46.5%
At or younger, below	24.4%	26.6%	28.8%
Older, at or above	27.0%	29.0%	24.7%
Race/ethnicity category			
White	66.0%	67.3%	69.3%
Black	13.7%	14.0%	13.8%
Hispanic	13.8%	12.4%	11.0%
Other	6.4%	6.2%	5.9%
Gender			
Male	52.4%	52.2%	52.3%
Female	47.6%	47.8%	47.7%
IEP			
Yes	6.0%	6.0%	4.7%
No	94.0%	94.0%	95.3%
LEP			
Yes	2.6%	1.6%	1.8%
No	97.4%	98.4%	98.2%
IEP, LEP			
IEP yes, LEP yes	0.1%	0.1%	0.1%
IEP yes, LEP no	5.9%	5.9%	4.6%
IEP no, LEP yes	2.5%	1.5%	1.8%
IEP no, LEP no	91.5%	92.5%	93.5%

Table 10-6

Distribution of Populations of Eligible Students Based on Invited Students from Participating Schools Using Weights Adjusted for School Nonresponse But Not Student Nonresponse 1994 Main NAEP Reading Sample

Population	Age Class 9	Age Class 13	Age Class 17
Total population	4,514,692	4,518,079	3,328,622
Age category			
At modal age or younger	73.0%	71.2%	75.2%
Older than modal age	27.0%	28.8%	24.8%
Grade category			
At or above modal grade	75.7%	73.6%	71.2%
Below modal grade	24.3%	26.4%	28.8%
Age category, grade category			
At or younger, at or above	48.6%	44.8%	46.5%
At or younger, below	24.3%	26.4%	28.8%
Older, at or above	27.0%	28.8%	24.8%
Race/ethnicity category			
White	65.7%	66.8%	69.0%
Black	13.5%	13.7%	13.6%
Hispanic	14.3%	13.2%	11.5%
Other	6.5%	6.3%	5.9%
Gender			
Male	52.3%	52.1%	52.2%
Female	47.7%	47.9%	47.8%
IEP			
Yes	6.0%	5.9%	4.6%
No	94.0%	94.1%	95.4%
LEP			
Yes	2.8%	1.8%	2.0%
No	97.2%	98.2%	98.0%
IEP, LEP			
IEP yes, LEP yes	0.1%	0.1%	0.1%
IEP yes, LEP no	5.9%	5.8%	4.5%
IEP no, LEP yes	2.7%	1.7%	2.0%
IEP no, LEP no	91.3%	92.4%	93.5%

Table 10-7

Distribution of Populations of Eligible Students Based on Assessed Students from Participating Schools with Student Nonresponse Adjustments Applied 1994 Main NAEP Reading Sample

Population	Age Class 9	Age Class 13	Age Class 17
Total population*	4,514,687	4,518,078	3,328,632
Age category			
At modal age or younger	72.7%	71.1%	75.0%
Older than modal age	27.3%	28.9%	25.0%
Grade category			
At or above modal grade	76.2%	73.7%	71.4%
Below modal grade	23.8%	26.3%	28.6%
Age category, grade category			
At or younger, at or above	48.8%	44.9%	46.3%
At or younger, below	23.8%	26.3%	28.6%
Older, at or above	27.3%	28.9%	25.0%
Race/ethnicity category			
White	65.7%	66.7%	68.8%
Black	12.9%	13.3%	13.1%
Hispanic	14.7%	13.5%	11.8%
Other	6.7%	6.5%	6.3%
Gender			
Male	52.3%	52.1%	52.0%
Female	47.7%	47.9%	48.0%
IEP			
Yes	5.9%	5.7%	4.2%
No	94.1%	94.3%	95.8%
LEP			
Yes	2.8%	1.8%	2.0%
No	97.2%	98.2%	98.0%
IEP, LEP			
IEP yes, LEP yes	0.1%	0.1%	0.1%
IEP yes, LEP no	5.8%	5.6%	4.2%
IEP no, LEP yes	2.7%	1.7%	1.9%
IEP no, LEP no	91.5%	92.6%	93.8%

******Note*: *Figures do not match those in Table 10-6 due to rounding.*

Table 10-8
Weighted Distribution of Absent Students by Nature of Absenteeism for All Age Classes
1994 Reading Assessment

Nature of Absenteeism	Age Class 9	Age Class 13	Age Class 17
Temporary absence*	69.4%	79.8%	68.9%
Long-term absence**	2.0%	1.9%	1.5%
Chronic truant	0.4%	2.6%	3.2%
Suspended or expelled	0.3%	3.0%	0.9%
Parent refusal	11.1%	2.9%	1.0%
Student refusal	0.0%	0.2%	9.6%
In school, did not attend session	1.1%	4.1%	9.8%
In school, not invited***	1.4%	1.2%	0.0%
Other	14.0%	4.3%	5.0%
Missing	0.0%	0.0%	0.0%
Total absentee sample	706	2,034	4,502
Total sample size	10,386	22,907	24,022
Overall absentee rate	6.8%	8.9%	18.7%

__Note__: Absent less than two weeks due to illness, disability, or excused absence.

**__Note__: Absent more than two weeks due to illness or disability.*

***__Note__: In school, but not invited to assessment session due to disruptive behavior.*

10.2.3 Trimming of Weights

In a number of cases, students were assigned relatively large weights. One cause of large weights was underestimation of the number of eligible students in some schools leading to inappropriately low probabilities of selection for those schools. A second major cause is the presence of large schools (high schools in particular) in PSUs with small selection probabilities. In such cases, the maximum permissible within-school sampling rate (determined by the maximum sample size allowed per school—see Chapter 3) could well be smaller than the desired overall within-PSU sampling rate for students. Large weights arose also because very small schools were, by design, sampled with low probabilities. Other large weights arose as the result of high levels of nonresponse coupled with low to moderate probabilities of selection, and the compounding of nonresponse adjustments at various levels.

Students with notably large weights have an unusually large impact on estimates such as weighted means. Since, under some simplifying assumptions, the variability in weights contributes to the variance of an overall estimate by an approximate factor $1 + V^2$, where V^2 is the relative variance of the weights, an occasional unusually large weight is likely to produce large sampling variances of the statistics of interest, especially when the large weights are associated with students with atypical performance characteristics.

To reduce this problem, a procedure of trimming a few of the more extreme weights to values somewhat closer to the mean weight was applied. This trimming can increase the accuracy of the resulting survey estimates, substantially reducing V^2 and hence the sampling variance while introducing a small bias. The trimming algorithm was identical to that used since 1984, and had the effect of trimming the weights of students from any school that contributed more than a specified proportion, ζ, to the estimated variance of the estimated number of students eligible for assessment. The trimming was done separately for each assessment. In each case, the value of the proportion ζ was chosen to be 10/K, where K was the number of schools in which a specified assessment was conducted. The number of schools where weights were trimmed was no more than seven in any one assessment. The most extreme trimming factors applied were of the order of 0.4 (except for one that was 0.3); trimming affects the weights of only a very small proportion of the assessed students.

Table 10-9 shows the distributions of eligible students based on the trimmed weights of assessed students for the reading samples for each age class. The distributions are similar to those before trimming shown in Tables 10-5, 10-6, and 10-7. To the extent that the characteristics in the table are related to student performance on the reading assessment, the bias in the assessment results introduced by trimming is very small.

Table 10-9

Distribution of Populations of Eligible Students Based on Trimmed Weights of Assessed Students in Participating Schools, 1994 Main NAEP Reading Sample

Population	Age Class 9	Age Class 13	Age Class 17
Total population	4,475,485	4,474,197	3,324,690
Age category			
At modal age or younger	72.7%	71.1%	75.0%
Older than modal age	27.3%	28.9%	25.0%
Grade category			
At or above modal grade	76.1%	73.8%	71.4%
Below modal grade	23.9%	26.2%	28.6%
Age category, grade category			
At or younger, at or above	48.8%	44.9%	46.3%
At or younger, below	23.9%	26.2%	28.6%
Older, at or above	27.3%	28.9%	25.0%
Race/ethnicity category			
White	65.6%	66.5%	68.8%
Black	13.0%	13.4%	13.1%
Hispanic	14.8%	13.6%	11.9%
Other	6.6%	6.5%	6.3%
Sex			
Male	52.4%	52.1%	52.0%
Female	47.6%	47.9%	48.0%
IEP			
Yes	5.9%	5.7%	4.2%
No	94.1%	94.3%	95.8%
LEP			
Yes	2.8%	1.8%	2.0%
No	97.2%	98.2%	98.0%
IEP, LEP			
IEP yes, LEP yes	0.1%	0.1%	0.1%
IEP yes, LEP no	5.8%	5.6%	4.2%
IEP no, LEP yes	2.7%	1.7%	1.9%
IEP no, LEP no	91.5%	92.6%	93.8%

10.2.4 Poststratification

As in most sample surveys, the respondent weights are random variables that are subject to sampling variability. Even if there were no nonresponse, the respondent weights would at best provide unbiased estimates of the various subgroup proportions. However, since unbiasedness refers to average performance over a conceptually infinite number of replications of the sampling, it is unlikely that any given estimate, based on the achieved sample, will exactly equal the population value. Furthermore, the respondent weights have been adjusted for nonresponse and a few extreme weights have been reduced in size.

To reduce the mean squared error of estimates using the sampling weights, these weights were further adjusted so that estimated population totals for a number of specified subgroups of the population, based on the sum of weights of students of the specified type, were the same as presumably better estimates based on composites of estimates from the 1990 and 1991 Current Population Survey and 1994 population projections made by the Census Bureau. This adjustment, called poststratification, is intended especially to reduce the mean squared error of estimates relating to student populations that span several subgroups of the population, and thus to reduce the variance of measures of changes over time for such student populations.

10.2.4.1 1994 Poststratification Procedures

The poststratification in 1994 was done for each age class and separately for each of the print-administered assessments and each of the groups of tape-administered assessments. Within each age class and assessment type group, poststratification adjustment cells were defined in terms of race, ethnicity, and NAEP region as shown in Table 10-10.

Table 10-10
Major Subgroups for Poststratification in 1994

Subgroup	Race	Ethnicity	Region*
1	White	Not Hispanic	Northeast
2	White	Not Hispanic	Southeast
3	White	Not Hispanic	Central
4	White	Not Hispanic	West
5	Any	Hispanic	Any
6	Black	Not Hispanic	Any
7	Other	Not Hispanic	Any

**Note*: Regions are the same as for stratification and reporting (see Chapter 3), except that all of Virginia is included in the Southeast region for poststratification purposes.*

The result is seven poststratification cells for each group of tape-administered sessions. Each of these cells accounts for between 4.6 percent (Subgroup 7) and 20.6 percent (Subgroup 3) of the population. For the assessments involving both age- and grade-eligible students, each of the seven subgroups was further divided into two or three eligibility classes. For age classes 9 and 13 and for the age 17/grade 11 long-term trend sample, three eligibility classes were used:

a) students eligible by grade and of modal age;
b) students eligible by age only; and
c) students eligible by grade but not of modal age.

For the age 17/grade 12 main assessment sample, each of the seven subgroups was divided into two subclasses:

a) students eligible by grade (of any age) and
b) students eligible by age only.

This variation in the procedure from that used for the other age classes and for the age 17/grade 11 long-term trend sample was adopted because the independent estimates of the numbers of students in the population did not provide consistent data on the numbers of twelfth grade students eligible only by grade. These poststratification cells have been used since 1988. (See Rust, Bethel, Burke, & Hansen, 1990, and Rust, Burke, & Fahimi, 1992, for further details.)

Thus, there were 7, 14, or 21 cells for poststratification. The poststratified weight for each student within a particular cell was the student's base weight, with adjustments for nonresponse and trimming, times a poststratification factor. For each cell, the poststratification factor is a ratio whose denominator is the sum of the weights (after adjustments for nonresponse and trimming) of assessed and excluded students, and whose numerator is an adjusted estimate of the total number of students in the population who are members of the cell. This estimated total was a composite based on the October 1990 and 1991 Current Population Surveys and 1994 population projections. Details of the procedures used to obtain these composite independent estimates are provided in Wallace and Rust (1996).

10.2.5 The Final Student Weight

The final weight assigned to a student is the student full-sample weight. This weight is the student's base weight after the application of the various adjustments described above. The student full-sample weight was used to derive all estimates of population and subpopulation characteristics that have been presented in the various NAEP reports, including simple estimates such as the proportion of students of a specified type who would respond in a certain way to an item and more complex estimates such as mean proficiency levels. The distributions of the final student weights are given in Tables 10-11 and 10-12.

As indicated earlier, under some simplifying assumptions the factor $1 + V^2$ indicates the approximate relative increase in variance of estimates resulting from the variability in the weights. The factor $1 + V^2$ for each sample is readily derivable from Tables 10-11 and 10-12 by adding 1 to the square of the ratio of the standard deviation to the mean weight. These factors, resulting from the combined effect of the variations in weights introduced by design and from other causes, are discussed in Section 10.2.3.

Table 10-11

Distribution of Final Student Weights, Long-term Trend Samples

Sample	Number of Cases	Mean	Standard Deviation	Minimum	25th Percentile	Median	75th Percentile	Maximum
AGE 9 LONG-TERM TREND SAMPLES								
Booklets 51-56	5335	915.7	472.6	167.5	557.8	863.3	1209.2	6662.5
Booklets 91-93	5663	584.3	261.4	151.2	409.7	531.8	671.8	2480.6
Excluded students	984	453.7	318.3	89.2	232.3	400.9	535.1	2291.4
AGE 13 LONG-TERM TREND SAMPLES								
Booklets 51-56	5547	884.3	381.9	223.9	604.7	851.3	1077.1	4288.4
Booklets 91-93	6052	554.1	249.7	108.8	429.8	491.6	612.2	2520.7
Excluded students	806	474.8	365.2	70.0	275.2	374.8	500.2	4252.4
AGE 17 LONG-TERM TREND SAMPLES								
Booklets 51-56	4840	878.5	482.7	144.9	531.3	722.6	1074.8	5596.9
Booklets 84-85	3813	807.4	338.7	243.9	555.3	697.2	1075.0	2175.4
Excluded students	504	552.5	316.3	151.0	346.5	488.6	656.5	2194.8

Table 10-12

Distribution of Final Student Weights, Main Samples

Sample	Number of Cases	Mean	Standard Deviation	Minimum	25th Percentile	Median	75th Percentile	Maximum
AGE 9/GRADE 4 MAIN SAMPLES								
Booklets 1-16 (Reading)	9680	515.6	283.5	92.1	297.3	429.4	668.8	2299.1
Booklets 31-48 (Geography)	7205	693.2	412.7	125.8	414.3	569.8	886.0	4632.2
Booklets 101-118 (U.S. History)	7205	692.8	414.9	117.3	400.1	568.3	889.8	4298.2
Excluded students	1512	224.8	113.1	39.0	140.4	193.7	285.6	1138.5
AGE 13/GRADE 8 MAIN SAMPLES								
Booklets 1-21 (Reading)	20873	239.7	146.1	33.8	136.1	202.0	303.0	3291.6
Booklets 31-49 (Geography)	9186	544.1	337.1	94.1	321.8	473.4	662.8	6017.0
Booklets 101-133 (U.S. History)	11669	428.6	267.2	57.1	246.5	360.3	531.5	5213.7
Excluded students	2044	141.6	84.9	44.5	89.2	121.1	174.2	1312.0
AGE 17/GRADE 12 MAIN SAMPLES								
Booklets 1-22 (Reading)	19520	194.2	106.2	21.9	114.3	163.7	250.5	712.2
Booklets 31-49 (Geography)	8589	435.7	230.4	9.5	260.3	368.4	572.6	2099.3
Booklets 101-133 (U.S. History)	10837	345.9	182.3	8.8	213.2	292.3	449.2	1660.9
Excluded students	1874	94.0	49.6	2.4	58.5	81.6	119.6	342.1

10.2.6 Other Weights

Weights were derived for excluded students as well as assessed students. In addition, special weighting adjustments were developed for certain subsets of the fourth-grade and eighth-grade students assessed in the main samples. The weights for these subsets, with these adjustments applied, were used in equating the results of the national assessment and the Trial State Assessments for subjects they had in common. Finally, a set of weights appropriate for analyzing school-level data files was developed.

Weights for Excluded Students. Excluded students are students who were designated by the schools as unable to complete the assessment because they were non-English speaking, mildly mentally retarded (educable), or functionally disabled. Since the same grade and age eligibility definitions apply, no distinction is made between students excluded from the various assessments within the long-term trend or main samples. However, the excluded students from the long-term trend and main assessments were treated as two separate samples of excluded students for each age class. This was in part because the guidelines to school personnel for excluding students differed between the main and long-term trend samples, so that the excluded student populations may have differed between these assessment types. The distribution of the final weights for excluded students are given in Tables 10-11 and 10-12.

For the long-term trend samples, students could be potentially excluded from a tape-administered session for which only age-eligible students were selected, or a print-administered session, for which both grade- and age- eligible students were selected. The samples of excluded students from the long-term trend assessments were weighted to reflect the full grade- and/or age-eligible population. This was achieved by weighting each grade-only eligible (i.e., not eligible by age) student who was excluded from a print-administered session to account for his or her probability of assignment to a print-administered session. No such corresponding session assignment adjustment was needed for the age-eligible excluded students, since they were eligible to have been selected for any of the long-term trend assessment.

As in the case of the weights for the assessed students, the excluded student weights were constructed from components reflecting the probability of selection, with correction for nonresponse, weight trimming, and poststratification. Further details on the derivation of the excluded student weights can be found in Wallace and Rust (1996).

Weights for Equating National and State-by-State Assessments. The fourth-grade reading assessments conducted in February 1994 as part of the Trial State Assessment Program consisted of identical assessment material to that administered in the corresponding national main sample sessions. Technical details of the Trial State Assessment Program are given in Mazzeo, Allen, and Kline (1995). The national and state-by-state assessments were equated so that state and national results could be reported on a common scale. The equating was achieved by using from each assessment that part of the sample representing a common population. For the age class 9 main sample, this consisted of those fourth-grade public-school students from a participating state (including the District of Columbia) who were assessed in the main sample reading assessment.

Although each sample of students received appropriate weights from the weighting procedure used for the national assessment, in an effort to increase the precision of the equating process, an additional weighting adjustment was developed and applied to this subsample, solely for use in equating. The distributions of the main sample weights for three categorical variables were adjusted to agree closely with those obtained from the weighted aggregate sample from the state assessments from the participating states. The first two variables were region (Northeast, Southeast, Central, and West) and race/ethnicity (White nonHispanic, Black nonHispanic, Hispanic, and "other"). The third variable was kind of reader (very good, good, average, poor) for fourth-grade reading. The equating of the weight distribution was achieved using a procedure known as iterative proportional fitting (described by Little & Rubin, 1987). Multiplicative adjustments were applied to the main sample weights to force their distribution to agree with that from the aggregated state samples, for each of these three variables in turn. This process was then repeated, and the final set of adjusted weights was compared with the state sample weights on all three distributions, and found to be in very close agreement. The resulting adjustments to the national weights ranged in magnitude from a factor of 0.65 to 1.42.

School Weights. The sampling procedures used to obtain national probability samples of assessed students also gave rise indirectly to several national probability samples of schools (from which the students were subsequently sampled). So that the school samples can be utilized for making national estimates about schools, appropriate nonresponse adjusted survey weights have been developed.

The weight for each school is partly composed of a base weight, giving the inverse of the selection probability of the school. This weight, W_{BS}, is given by

$$W_{BS} = PSUWT \cdot SCHWT.$$

School nonresponse adjustments were then applied to these base weights. These are very similar to the school nonresponse adjustment factors used for student weights and were created using the same set of nonresponse adjustment classes. The values of the adjustment factors are not the same, however. A school that was assigned a given assessment but did not participate at all was treated as not responding for the student weighting for that assessment (since its nonparticipation did not affect those session types that were not assigned to it). Such a school was treated as a nonresponding school in creating the school weights.

Six samples of selected schools were weighted to be nationally representative. At each age class, there were two such samples, one for the long-term trend assessment, and one for the main assessment. At age class 9, the population of schools represented in each case consists of all schools having at least one of the grades 2 through 5. The school population at age class 13 is that of schools having at least one of the grades 6 through 9, while the school population at age class 17 is that of schools having at least one of the grades 9 through 12.

Jackknife Replicate Weights. In addition to the weights that were used to derive all estimates of population and subpopulation characteristics, other sets of weights, called jackknife replicate weights, were derived to facilitate the estimation of sampling variability by the jackknife variance estimation technique. These weights and the jackknife estimator are discussed in the next section.

10.3 PROCEDURES USED BY NAEP TO ESTIMATE SAMPLING VARIABILITY

A major source of uncertainty in the estimation of the value in the population of a variable of interest exists because information about the variable is obtained on only a sample from the population. To reflect this fact, it is important to attach to any statistic (e.g., a mean) an estimate of the sampling variability to be expected for that statistic. Estimates of sampling variability provide information about how much the value of a given statistic would be likely to change if the statistic had been based on another, equivalent, sample of individuals drawn in exactly the same manner as the achieved sample.

Another important source of variability is that due to imprecision in the measurement of individual proficiencies. For the 1994 assessment, proficiencies in all subject areas were summarized through item response theory (IRT) models, but not in the way that these models are used in standard applications where each person responds to enough items to allow for precise estimation of that person's proficiency. In NAEP, each individual responds to relatively few items so that individual proficiency values are not well determined. Consequently, the variance of any statistic based on proficiency values has a component due to the imprecision in the measurement of the proficiencies of the sampled individuals in addition to a component measuring sampling variability. The estimation of the component of variability due to measurement imprecision and its effect on the total variability of statistics based on proficiency values are discussed in Chapter 11.

The estimation of the sampling variability of any statistic must take into account the sample design. In particular, because of the effects of cluster selection (students within schools, schools within PSUs) and because of effects of nonresponse and poststratification adjustments, observations made on different students cannot be assumed to be independent of each other (and are, in fact, generally positively correlated). Furthermore, to account for the differential probabilities of selection (and the various adjustments), each student has an associated sampling weight, which should be used in the computation of any statistic and which is itself subject to sampling variability. Ignoring the special characteristics of the sample design and treating the data as if the observations were independent and identically distributed, will generally produce underestimates of the true sampling variability, due to the clustering and unequal sampling weights.

The proper estimation of the sampling variability of a statistic based on the NAEP data is complicated and requires techniques beyond those commonly available in standard statistical packages. Fortunately, the *jackknife* procedure (see, e.g., Wolter, 1985; Kish & Frankel, 1974; Rust, 1985) provides good quality estimates of the sampling variability of most statistics, at the expense of increased computation, and can be used in concert with standard statistical packages to obtain a proper estimate of sampling variability.

The jackknife procedure used by NAEP has a number of properties that make it particularly suited for the analysis of NAEP data. When properly applied, a jackknife estimate of the variability of a linear estimator (such as a total) will be the same as the standard textbook variance estimate specified for the sample design (if the first-stage units were sampled with replacement and approximately so otherwise). Additionally, if the finite sampling corrections for the first stage units can be ignored, the jackknife produces asymptotically consistent variance estimates for statistics such as ratios, regression estimates or weighted means and for any other nonlinear statistic that can be expressed as a smooth function of estimated totals of one or more variables (Krewski & Rao, 1981).

Through the creation of student replicate weights (defined below), the jackknife procedure allows the measurement of variability attributable to the use of poststratification and other weight adjustment factors that are dependent upon the observed sample data. Once these replicate weights are derived, it is a straightforward matter to obtain the jackknife variance estimate of any statistic.

The jackknife procedure in this application is based upon the development of a set of jackknife replicate weights for each assessed student (or excluded student, or school depending upon the file involved). The replicate weights are developed in such a way that, when utilized as described below, approximately unbiased estimates of the sampling variance of an estimate result, with an adequate number of degrees of freedom to be useful for purposes of making inferences about the parameter of interest.

The estimated sampling variance of a parameter estimator t is the sum of M squared differences (where M is the number of replicated weights developed):

$$\hat{Var}(t) = \sum_{i=1}^{M} (t_i - t)^2$$

where t_i denotes the estimator of the parameter of interest, obtained using the ith set of replicate weights, $SRWT_i$, in place of the original set of full sample estimates WT.

In the main assessment samples, 62 replicate weights were developed using the procedures outlined below. Similar procedures were followed for the long term trend samples. However, since those samples were based on fewer PSUs (52 rather than 94), the long term trend samples have fewer replicate weights (36 instead of 62). Full details of the generation of replicate weights for all samples are given in Wallace and Rust (1996).

Of the 62 replicate weights formed for each record from a main assessment sample, 36 act to reflect the amount of sampling variance contributed by the noncertainty strata of PSUs, with the remaining 26 replicate weights reflecting the variance contribution of the certainty PSU samples.

The derivation of the 36 replicate weights reflecting the variance of the noncertainty PSUs involves first defining pairs of PSUs in a manner that models the design as one in which two PSUs are drawn with replacement per stratum. This definition of pairs is undertaken in a manner closely reflective of the actual design, in that PSUs are pairs that are drawn from strata within the same subuniverse, and with similar stratum characteristics. The same definition of pairs was used for each of the age/grade classes in the main assessment, since all were drawn from the same sample of noncertainty PSUs. The 72 noncertainty PSUs, drawn one from each of 72 strata, were formed into 36 pairs of PSUs, where the pairs were composed of PSUs from adjacent strata within each subuniverse (thus the strata were relatively similar on socioeconomic characteristics such as proportion minority population, population change since 1980, per capita income, educational attainment, and unemployment rate). Whereas the actual sample design was to select one PSU with probability proportional to size from each of 72 strata, for variance estimation purposes the design is regarded as calling for the selection of two PSUs with probability proportional to size with replacement from each of 36 strata. This procedure likely gives a small positive bias to estimates of sampling error.

The student replicate weight for the i^{th} pair of noncertainty PSUs, for the 36 pairs corresponding to values of i from 1 to 36, is computed as follows:

1) Let W_B be the base weight of a student, as described in Section 10.2.1, which accounts for the various components of the selection probability for the student.

2) At random, one PSU in each pair is denoted as PSU number 1, while the other is denoted as PSU number 2. The i^{th} replicate base weight, W_{Bi}, is given by:

$$W_{bi} = \begin{cases} 0 & \text{if the student belongs to PSU number 1 of pair i} \\ 2 * W_B & \text{if the student belongs to PSU number 2 of pair i} \\ W_B & \text{if the student is from neither PSU in pair i} \end{cases}$$

3) The i^{th} student replicate weight $SRWT_i$ is obtained by applying the various school and student nonresponse adjustments, the weight trimming, and the poststratification to the i^{th} set of replicate base weights, using procedures identical to those used to obtain the final student weights WT from the set of base weights W_B.

In brief, the procedure for deriving the sets of W_{Bi} values from the W_B values reflects the sampling of PSUs, schools, sessions, and students. By repeating the various weight adjustment procedures in each set of replicate base weights, the impact of these procedures on the sampling variance of the estimator t is appropriately reflected in the variance estimator $\hat{Var}(t)$ defined above.

The procedure for obtaining the 26 sets of replicate weights to estimate the sampling variance from the certainty PSUs is analogous, but somewhat more complex. The first stage of sampling in this case is at the school level, and the derivation of replicate weights must reflect appropriately the sampling of schools within certainty PSUs. Since each of the three age/grade classes in the main assessment involved different samples of schools, the procedure for forming replicate base weights was individualized to each of these sample components. In common across these three samples were the 22 certainty PSUs used, and the fact that 26 replicate weights were formed in each case.

For a given sample, the 22 certainty PSUs constituted strata, with a sample of schools drawn systematically within each. Using the schools listed in order of sample selection within each stratum, successive schools were paired or formed into triples. These pairs and triples numbered more than 26, so that each replicate weight was in general formed by perturbing the weights of students from more than a single pair or triple. These aggregates of pairs and triples were in general assigned in proportion to the size of the PSU. Thus generally speaking, the largest PSUs were assigned three replicates each, the next largest were assigned two replicates each, and the remaining self-representing PSUs were assigned one replicate each. When splitting the larger PSUs, the schools were split into groups of (as close as possible) equal size, based on the ordering at the time of sample selection. One group was assigned to each replicate. Within each PSU (or partial PSU in the case of the large split PSUs) schools were alternately numbered 1 or 2 starting randomly. If, however, there were exactly three schools sampled in the PSU the schools were randomly numbered 1, 2, or 3. The method of forming replicate base weights in strata where there were not exactly three schools was the same as for the noncertainty strata (except that members of a pair i could come from more than a single "stratum"). When a stratum contained three schools, students in these schools had their weights perturbed for two sets of replicates, say i1 and i2, as follows:

$$W_{bi1} = \begin{cases} 0 & \text{if the student is in school number 1 of a PSU belonging to set i} \\ 1.5 * W_B & \text{if the student is in school number 2 or 3 of a PSU belonging to set i} \\ W_B & \text{if the student does not belong to a PSU in set i} \end{cases}$$

$$W_{bi2} = \begin{cases} 1.5 * W_B & \text{if the student is in school number 1 or 2 of a PSU belonging to set i} \\ 0 & \text{if the student is in school number 3 of a PSU belonging to set i} \\ W_B & \text{if the student does not belong to a PSU in set i} \end{cases}$$

The actual pattern of replicate base weight assignment used for each of the samples is given in Wallace and Rust (1996).

The nonresponse, trimming, and poststratification adjustments were applied to each set of replicate base weights to derive the final replicate weights in each case, exactly as in the noncertainty PSUs. In fact these procedures were applied to the full set of weights from all parts of the given sample together, just as for the full sample weights. That is, for example, poststratification factors were derived from the full set of data for each replicate, not separately for certainty and noncertainty PSUs.

This estimation technique was used by NAEP to estimate all sampling errors presented in the various reports. A further discussion of the variance estimation procedure used by NAEP, including a discussion of alternative jackknife estimators that were also considered, appears in Johnson (1989).

We noted above (as discussed in Chapter 11) that a separate estimate of the contribution to variance due to the imprecision in the measure of individual proficiencies is made and added to the jackknife estimate of variance. That variance component could have been approximately reflected in the jackknife variance estimates simply by separately applying the IRT computations to each jackknife replicate. Because of the heavier IRT computational load, this was not done. Less work was involved by the simple procedure of making separate estimates of this component to be added to the jackknife variance estimates. Also, a separate measure of this component of variance is then available, which would not be so if it were reflected in the jackknife variance estimate.

10.4 APPROXIMATING THE SAMPLING VARIANCE USING DESIGN EFFECTS

In practical terms, the major expenditure of resources in the computation of a jackknife variance estimate occurs in the preparation of estimates for each of the pseudoreplicates. In the 1994 assessment, this implies that the statistic of interest has to be recomputed up to 63 times, once for the overall estimate t, and once for each of the up to 62 pseudoreplicates t_i. Because this is a considerable increase in the amount of computation required, relative to a conventional variance estimate, it is of interest to see how much the jackknife variance estimates differ from their less computationally intensive, simple random sampling based, analogues.

The comparison of the conventional and the jackknife methods of variance estimation will be in terms of a statistic called the *design effect*, which was developed by Kish (1965) and extended by Kish and Frankel (1974). The design effect for a statistic is the ratio of the actual variance of the statistic (taking the sample design into account) over the conventional variance estimate based on a simple random sample with the same number of elements. The design effect is the inflation factor to be applied to the conventional variance estimate in order to adjust error estimates based on simple random sampling assumptions to account approximately for the effect of the sample design. The value of the design effect depends on the type of statistic computed and the variables considered in a particular analysis as well as the combined clustering, stratification, and weighting effects occurring among sampled elements. While stratification drives down the sampling variance, the effects of clustering and weighting that drive variances up are generally sufficient to produce variance estimates that are larger than variances based on simple random sampling assumptions. Consequently, the design effects will be greater than one. In NAEP, the underestimates are the result of ignoring the effects of clustering and unequal probabilities of selection in the variance calculations.

Since most of the analyses conducted by NAEP are based on the results of scaling models that summarize performance of students across a learning area, we consider the design effects to be expected for analyses based on these scale scores. For reasons given in Chapter 11, NAEP provides each individual with a set of "plausible values," each of which is a random draw from the distribution of the potential scale scores for that individual. Since our current interest is on the effect of the sampling design on estimation and inference, we will restrict our attention to a single measure of an individual's proficiency, the first plausible value of the individual's scale score.

A key statistic of interest is the estimated mean proficiency of a subgroup of the population. An estimate of the subgroup mean proficiency is the weighted mean of the first plausible values of proficiency of the sampled individuals who belong to the subpopulation of interest. Let Y be the weighted mean of the plausible values of the sampled members of the subpopulation. The conventional estimate of the variance of Y is

$$Var_{con}(\bar{Y}) = \frac{\sum_{i=1}^{N} w_i \, (y_i - \bar{Y})^2}{N \, W_{.}}$$

where N is the total number of sampled individuals in the subpopulation for which plausible values are available, w_i is the weight of the i^{th} individual, y_i is a plausible value from the distribution of potential proficiencies for that individual, and W_{+} is the sum of the weights across the N individuals.

The design effect for the subgroup mean proficiency estimate is

$$\text{deff}(\bar{Y}) = \text{Var}_{JK}(\bar{Y}) / \text{Var}_{con}(\bar{Y})$$

where $\text{Var}_{JK}(\bar{Y})$ is the jackknife variance of $\bar{Y}$. (As has been pointed out previously, $\text{Var}_{JK}(\bar{Y})$ as computed does not measure the variability of Y due to imprecision in the measurement of the proficiencies of the sampled individuals. The estimation of this very important source of variability is discussed in Chapter 11.)

Values of the design effects for subgroup mean proficiencies are displayed, by grade, in Tables 10-13, 10-14, and 10-15, for the main assessments of U.S. history, geography, and reading, respectively. Design effects are shown for the population as a whole (Total) as well as for a variety of demographic subgroups: gender; race/ethnicity (White, Black, Hispanic, Asian American, other); type of location (rural, suburban, urban); parental education (did not graduate high school, graduated high school, post-high school, graduated college, unknown); and type of school (public, nonpublic). These particular demographic variables were selected because (1) they are major variables in NAEP reports and (2) they reflect different types of divisions of the population that might have different levels of sampling variability.

Tables 10-16 and 10-17 provide equivalent information for the long-term trend samples. Table 10-16 provides, for each age class and demographic subgroup, the average of the design effects for reading and writing mean proficiencies for the students selected for the long-term trend assessments of reading and writing. Table 10-17 provides the average of the design effects for the mean mathematics and science proficiencies for the students selected for the long-term assessments of mathematics and science.

Finally, for comparison with the national reading results, Table 10-18 gives the average design effects for state-level mean reading proficiency, averaged across all jurisdictions participating in the grade 4 Trial State Assessment of reading.

The tables show that the design effects are predominantly larger than 1, indicating that standard variance estimation formulas will be generally too small, usually markedly so. Although the design effects appear somewhat different for certain subgroups of the population, they are, perhaps, similar enough (at least within a subject and grade) to select an overall composite value that is adequate for most purposes. In choosing a composite design effect, some consideration must be made about the relative consequences of overestimating the variance as opposed to underestimating the variance. For example, if an overestimate of the variance is viewed as severe an error as an underestimate, the composite design effect should be near to the center of the distributions of the design effects. Possible composites of this type are the mean and median design effects across the combined distribution of all design effects. Larger design effects should be used if it is felt that it is a graver error to underestimate the variability of a statistic than to overestimate it. For example, Johnson and King (1987) examine estimation of variances using design effects (among other techniques) under the assumption that the consequences of an underestimate are three times as severe as those of an overestimate of the same magnitude. Adopting a loss function that is a weighted sum of absolute values of the deviations of predicted from actual with underestimates receiving three times the weight of overestimates, produces the upper quartile of the design effects as the composite value. This assumes that the distribution of design effects is roughly independent of the jackknife estimates of variance, so that the size of a design effect does not depend on the size of the variance.

Table 10-19 gives the values of these potential composites, by grade, for the U.S. history, geography, and reading assessments, and across those assessments. Table 10-20 gives composite values for the Trial State Assessment of grade 4 reading. The state assessment tends to have smaller design effects than the matching national assessment, due to the lesser degree of clustering in the state assessment samples.

Table 10-21 gives the values of the composites for the two long-term trend samples. The mathematics and science long-term trend samples have larger design effects than do the reading and writing long-term trend samples. This is because the mathematics and science samples used paced-tape administration while the reading and writing samples used a spiraled, paper-and-pencil administration. Because the same items must be administered in a tape-administered session, the number of students per school administered an item is higher for samples with paced-tape administration than for print-administered (spiraled) samples of a similar size. This results in more clustering.

We note that the $\mathrm{Var}_{con}(\overline{Y})$ as defined above is an estimate of S^2/N where S^2 represents the unit variance for a simple random sample for the population of students from which the sample is also drawn. This is an appropriate estimate of the increase in variance over simple random sampling from that population. However, the computer packages used for estimating the variance may not reflect the weights in estimating the unit variance, as given above, but instead may provide an estimate of a unit variance of the form

$$\frac{\sum_{i=1}^{N} (y_i - \bar{Y})^2}{N^2}$$

In this case, the unweighted estimate of unit variance would be appropriate for the denominator of a design effect measure of the increase in variance over the unit variance as estimates by the computer package. If there is no correlation between the w_i and y_i, there would be little difference between the two.

Table 10-13

*Design Effects by Demographic Subgroup and Grade for Mean U.S. History Proficiencies**

Subgroup	Grade 4	Grade 8	Grade 12
Total	3.32	3.17	4.25
Male	2.89	2.28	2.54
Female	2.12	2.02	2.68
White	3.41	3.57	3.72
Black	1.41	2.78	2.58
Hispanic	2.94	2.13	1.49
Asian American	1.57	5.76	3.35
Other Race/ethnicity	2.11	1.59	1.12
Urban	4.64	5.08	4.08
Suburban	3.77	4.72	3.81
Rural	3.35	3.09	1.92
Par. Ed. < HS	1.19	1.36	1.05
Par. Ed. = HS	1.28	1.36	1.44
Par. Ed. > HS	1.22	1.15	2.90
Par. Ed. = College	2.16	2.11	2.36
Par. Ed. = IDK	2.00	1.39	1.03
Public School	3.48	3.36	4.24
Nonpublic School	3.16	2.23	2.76

**Note: Design effects are based on the conventional and jackknife variances of subgroup means of the first plausible values of proficiency.*

Table 10-14

*Design Effects by Demographic Subgroup and Grade for Mean Geography Proficiencies**

Subgroup	Grade 4	Grade 8	Grade 12
Total	4.11	2.77	3.65
Male	2.74	2.15	2.31
Female	2.91	1.80	2.50
White	5.29	2.79	3.34
Black	3.14	3.28	2.28
Hispanic	2.25	2.67	1.73
Asian American	1.84	6.22	2.35
Other Race/ethnicity	0.82	1.71	0.57
Urban	4.68	4.99	4.02
Suburban	5.16	3.40	4.85
Rural	3.24	4.62	2.15
Par. Ed. < HS	1.33	1.42	0.90
Par. Ed. = HS	1.90	1.90	1.62
Par. Ed. > HS	1.12	1.49	2.20
Par. Ed. = College	2.82	2.26	2.60
Par. Ed. = IDK	2.00	1.41	0.94
Public School	4.09	2.81	3.33
Nonpublic School	3.94	2.46	4.54

**__Note__: Design effects are based on the conventional and jackknife variances of subgroup means of the first plausible values of proficiency.*

Table 10-15

*Design Effects by Demographic Subgroup and Grade for Mean Reading Proficiencies**

Subgroup	Grade 4	Grade 8	Grade 12
Total	3.82	4.84	2.75
Male	2.79	3.45	2.33
Female	2.56	3.19	1.97
White	3.86	4.84	2.25
Black	1.87	3.54	2.09
Hispanic	3.86	1.79	1.50
Asian American	3.81	3.94	1.23
Other Race/ethnicity	1.36	1.56	1.74
Urban	7.39	6.37	3.04
Suburban	5.92	4.77	3.23
Rural	2.25	4.74	3.28
Par. Ed. < HS	1.61	1.85	1.21
Par. Ed. = HS	1.51	2.10	2.30
Par. Ed. > HS	1.54	2.61	1.08
Par. Ed. = College	2.69	2.67	1.77
Par. Ed. = IDK	1.37	1.29	1.42
Public School	3.72	4.04	2.17
Nonpublic School	6.89	3.28	8.13

****Note****: Design effects are based on the conventional and jackknife variances of subgroup means of the first plausible values of proficiency.*

BEST COPY AVAILABLE

Table 10-16

*Design Effects by Demographic Subgroup and Age Averaged Over Mean Reading and Writing Proficiencies for the Reading and Writing Long-Term Trend Samples**

Subgroup	Age 9	Age 13	Age 17
Total	1.74	1.74	2.95
Male	1.68	1.81	2.34
Female	1.27	1.20	2.52
White	1.69	1.61	3.18
Black	1.13	0.96	2.21
Hispanic	2.49	1.11	2.58
Asian American	0.79	2.39	1.92
Other Race/ethnicity	0.99	1.36	1.24
Urban	5.24	2.97	3.61
Suburban	2.17	3.26	3.85
Rural	2.88	2.54	5.09
Par. Ed. < HS	1.54	0.96	1.67
Par. Ed. = HS	1.11	1.52	1.29
Par. Ed. > HS	1.12	1.47	1.38
Par. Ed. = College	1.12	1.35	2.51
Par. Ed. = IDK	1.44	0.80	0.98
Public School	2.38	1.70	3.27
Nonpublic School	1.80	1.64	3.49

**Note: Design effects are based on the conventional and jackknife variances of subgroup means of the first plausible values of proficiency.*

Table 10-17

*Design Effects by Demographic Subgroup and Age Averaged Over Mean Mathematics and Science Proficiencies for the Mathematics and Science Long-Term Trend Samples**

Subgroup	Age 9	Age 13	Age 17
Total	2.84	4.20	4.07
Male	2.38	2.72	2.73
Female	1.50	3.22	2.95
White	2.26	3.03	3.67
Black	0.78	9.52	2.70
Hispanic	3.70	2.91	5.03
Asian American	1.39	2.17	3.09
Other Race/ethnicity	1.27	1.27	1.16
Urban	7.19	6.96	7.65
Suburban	3.79	4.04	3.76
Rural	2.34	5.70	2.80
Par. Ed. < HS	0.91	1.50	1.72
Par. Ed. = HS	0.77	1.35	1.10
Par. Ed. > HS	0.89	2.07	1.65
Par. Ed. = College	2.03	2.96	2.48
Par. Ed. = IDK	2.39	1.96	1.30
Public School	3.71	4.42	3.17
Nonpublic School	3.31	3.74	6.37

****Note**: Design effects are based on the conventional and jackknife variances of subgroup means of the first plausible values of proficiency.*

Table 10-18

*Average Design Effects by Demographic Subgroup for Mean State Reading Proficiencies Averaged Across State Samples**

Subgroup	Grade 4
Total	3.70
Male	2.44
Female	2.45
White	2.57
Black	1.99
Hispanic	1.86
Asian American	1.46
Other Race/ethnicity	1.45
Urban	5.32
Suburban	4.42
Rural	3.49
Par. Ed. < HS	1.19
Par. Ed. = HS	1.41
Par. Ed. > HS	1.27
Par. Ed. = College	2.52
Par. Ed. = IDK	1.93
Public School	3.45
Nonpublic School	4.79

**Note:* *Design effects are based on the conventional and jackknife variances of subgroup means of the first plausible values of proficiency.*

Table 10-19

Within-Grade Mean, Median, and Upper Quartile of the Distribution of Design Effects for National Main Assessments by Subject Area and Across Subject Areas

Statistic	Grade 4	Grade 8	Grade 12
Mean U.S. History Proficiencies (Distribution Across Demographic Subgroups)			
Upper Quartile	3.35	3.36	3.72
Mean	2.56	2.73	2.63
Median	2.53	2.26	2.63
Mean Geography Proficiencies (Distribution Across Demographic Subgroups)			
Upper Quartile	4.09	3.28	3.34
Mean	2.97	2.79	2.55
Median	2.87	2.57	2.33
Mean Reading Proficiencies (Distribution Across Demographic Subgroups)			
Upper Quartile	3.86	4.74	2.75
Mean	3.27	3.38	2.42
Median	2.74	3.37	2.13
Across Subject Areas (Distribution Across Subject Areas and Demographic Subgroups)			
Upper Quartile	3.81	3.57	3.28
Mean	2.93	2.97	2.53
Median	2.81	2.72	2.32

Table 10-20

Mean, Median, and Upper Quartile of the Across-State Average Design Effects for Mean State Reading Proficiency (Distribution Across Demographic Subgroups)

Statistic	Grade 4
Upper Quartile	3.49
Mean	2.65
Median	2.45

Table 10-21

*Mean, Median, and Upper Quartile of the Distribution of Design Effects for the Long-Term Trend Assessment Samples**

Statistic	Age 9	Age 13	Age 17
Reading and Writing Long-Term Trend (Distribution Across Demographic Subgroups of Average of Design Effects for Reading and Writing Mean Proficiencies)			
Upper Quartile	2.17	1.81	3.27
Mean	1.81	1.69	2.56
Median	1.61	1.57	2.52
Mathematics and Science Long-Term Trend (Distribution Across Demographic Subgroups of Average of Design Effects for Mathematics and Science Mean Proficiencies)			
Upper Quartile	3.31	4.20	3.76
Mean	2.41	3.54	3.19
Median	2.30	3.00	2.88

*Note: *Design effects are based on the conventional and jackknife variances of subgroup means of the first plausible values of proficiency.*

Chapter 11

SCALING PROCEDURES[1]

Nancy L. Allen, Eugene G. Johnson, Robert J. Mislevy, and Neal Thomas
Educational Testing Service

11.1 INTRODUCTION

The primary method by which results from the 1994 National Assessment of Educational Progress (NAEP) were disseminated was scale-score reporting. With scaling methods, the performance of a sample of students in a subject area or subarea can be summarized on a single scale or series of scales even when different students have been administered different items. This chapter presents an overview of the scaling methodologies employed in the analyses of the data from NAEP surveys in general and from the 1994 assessment in particular. Details of the scaling procedures specific to the subject areas of reading, U.S. history, geography, mathematics, science, and writing are presented in Chapters 12 through 18.

11.2 BACKGROUND

The basic information from an assessment consists of the responses of students to the items presented in the assessment. For NAEP, these items are constructed to measure performance on sets of objectives developed by nationally representative panels of learning area specialists, educators, and concerned citizens. Satisfying the objectives of the assessment and ensuring that the tasks selected to measure each goal cover a range of difficulty levels typically require many items. For example, the reading assessment required 122 items at grade 8. Depending on the subject areas, a mixture of multiple-choice, short constructed-response, and extended constructed-response items were used. Multiple-choice and short constructed-response items were used in all assessments but writing. Extended constructed-response items, scored on a multipoint scale, were presented in the main reading, U.S history, and geography assessments and in the long-term trend writing assessment. To reduce student burden, each assessed student was presented only a fraction of the full pool of items through multiple matrix sampling procedures.

The most direct manner of presenting the assessment results is to report separate statistics for each item. However, because of the vast amount of information, having separate results for each of the items in the assessment pool hinders the comparison of the general performance of subgroups of the population. Item-by-item reporting masks similarities in trends and subgroup comparisons that are common across items.

---

[1] Nancy L. Allen is responsible for the psychometric and statistical analysis of national and state NAEP data. Eugene G. Johnson is a senior psychometrician with special expertise in the design of NAEP and sampling issues. Robert J. Mislevy is a technical consultant with expertise in item response theory. Neal Thomas is a technical consultant with expertise in imputed values.

An obvious summary of performance across a collection of items is the average of the separate item scores. The advantage of averaging is that it tends to cancel out the effects of peculiarities in items that can affect item difficulty in unpredictable ways. Furthermore, averaging makes it possible to compare more easily the general performances of subpopulation .

Despite their advantages, there are a number of significant problems with average item scores. First, the interpretation of these results depends on the selection of the items; the selection of easy or difficult items could make student performance appear to be overly high or low. Second, the average score is related to the particular items comprising the average, so that direct comparisons in performance between subpopulations require that those subpopulations have been administered the same set of items. Third, because this approach limits comparisons to average scores on specific sets of items, it provides no simple way to report trends over time when the item pool changes. Finally, direct estimates of parameters or quantities such as the proportion of students who would achieve a certain score across the items in the pool are not possible when every student is administered only a fraction of the item pool. While the average score across all items in the pool can be readily obtained (as the average of the individual item scores), statistics that provide distributional information, such as quantiles of the distribution of scores across the full set of items, cannot be readily obtained without additional assumptions.

These limitations can be overcome by the use of response scaling methods. If several items require similar skills, the regularities observed in response patterns can often be exploited to characterize both respondents and items in terms of a relatively small number of variables. These variables include a respondent-specific variable, called proficiency, which quantifies a respondent's tendency to answer items correctly (or, for multipoint items, to achieve a certain score) and item-specific variables that indicate characteristics of the item such as its difficulty, effectiveness in distinguishing between individuals with different levels of proficiency, and the chances of a very low proficiency respondent correctly answering a multiple-choice item. (These variables are discussed in more detail in the next section.) When combined through appropriate mathematical formulas, these variables capture the dominant features of the data. Furthermore, all students can be placed on a common scale, even though none of the respondents takes all of the items within the pool. Using the common scale, it becomes possible to discuss distributions of proficiency in a population or subpopulation and to estimate the relationships between proficiency and background variables.

It is important to point out that any procedure of aggregation, from a simple average to a complex multidimensional scaling model, highlights certain patterns at the expense of other potentially interesting patterns that may reside within the data. Every item in a NAEP survey is of interest and can provide useful information about what young Americans know and can do. The choice of an aggregation procedure must be driven by a conception of just which patterns are salient for a particular purpose.

The scaling for the assessments in reading, U.S. history, and geography was carried out separately within the content areas specified in the frameworks for those subjects. This scaling within subareas was done because it was anticipated that different patterns of performance or different trends over time might exist for these essential subdivisions of the subject areas. By creating a separate scale for each of these content areas, potential differences in subpopulation performance between the content areas are preserved.

The creation of a series of separate scales to describe performance within a subject area does not preclude the reporting of a single index of overall performance in the subject area—that is, an overall subject area composite. A composite is computed as the weighted average of the content area scales, where the weights correspond to the relative importance given to each content area as defined by the framework. The composite provides a global measure of performance within the subject area, while the constituent content area scales allow the measurement of important interactions within educationally relevant subdivisions of the subject area.

11.3 SCALING METHODOLOGY

This section reviews the scaling models employed in the analyses of data from the 1994 assessment, and the multiple imputation or "plausible values" methodology that allows such models to be used with NAEP's sparse item-sampling design. The reader is referred to Mislevy (1991) for an introduction to plausible values methods and a comparison with standard psychometric analyses, to Mislevy, Johnson and Muraki (1992) and Beaton and Johnson (1992) for additional information on how the models are used in NAEP, and to Rubin (1987) for the theoretical underpinnings of the approach. It should be noted that the imputation procedure used by NAEP is a mechanism for providing plausible values for the unobserved proficiencies and not for filling in blank responses to background or cognitive variables.

While the NAEP procedures were developed explicitly to handle the characteristics of NAEP data, they build on other research, and are paralleled by other researchers. See, for example Dempster, Laird, and Rubin (1977); Little and Rubin (1983, 1987); Andersen (1980); Engelen (1987); Hoijtink (1991); Laird (1978); Lindsey, Clogg, and Grego (1991); Zwinderman (1991); Tanner and Wong (1987); and Rubin (1987, 1991).

11.3.1 The Scaling Models

Three distinct scaling models, depending on item type and scoring procedure, were used in the analysis of the data from the 1994 assessment. Each of the models is based on item response theory (IRT; e.g., Lord, 1980). Each is a "latent variable" model, defined separately for each of the scales, which expresses respondents' tendencies to achieve certain scores (such as correct/incorrect) on the items contributing to a scale as a function of a parameter that is not directly observed, called proficiency on the scale.

A three-parameter logistic (3PL) model was used for the multiple-choice items (which were scored correct/incorrect). The fundamental equation of the 3PL model is the probability that a person whose proficiency on scale k is characterized by the *unobservable* variable θ_k will respond correctly to item j:

$$\begin{aligned} P(x_j = 1 | \theta_k, a_j, b_j, c_j) &= c_j + \frac{(1 - c_j)}{1 + \exp[-1.7a_j(\theta_k - b_j)]} \\ &\equiv P_{j1}(\theta_k) \ , \end{aligned} \tag{11.1}$$

where

x_j is the response to item j, 1 if correct and 0 if not;

a_j where $a_j > 0$, is the slope parameter of item j, characterizing its sensitivity to proficiency;

b_j	is the threshold parameter of item *j*, characterizing its difficulty; and
c_j	where $0 \le c_j < 1$, is the lower asymptote parameter of item *j*, reflecting the chances of students of very low proficiency selecting the correct option.

Further define the probability of an incorrect response to the item as

$$P_{j0} \equiv P(x_j = 0 | \theta_k, a_j, b_j, c_j) = 1 - P_{j1}(\theta_k) \qquad (11.2)$$

A two-parameter logistic (2PL) model was used for the short constructed-response items that were scored correct or incorrect. The form of the 2PL model is the same as equations (11.1) and (11.2) with the c_j parameter fixed at zero.

In addition to the multiple-choice and short constructed-response items, a number of extended constructed-response items were presented in the assessments of reading, U.S. history, and geography; and only extended constructed-response items were presented in the long-term trend writing assessment. Each of these items was scored on a multipoint scale with potential scores ranging from 0 to 3 or from 0 to 4. Some short constructed-response items were scored on a three-point scale (0-2). Items that are scored on a multipoint scale are referred to as polytomous items, in contrast with the multiple-choice and short constructed-response items, which are scored correct or incorrect and referred to as dichotomous items.

The polytomous items were scaled using a generalized partial credit model (Muraki, 1992). The fundamental equation of this model is the probability that a person with proficiency θ_k on scale *k* will have, for the *j*th item, a response x_j that is scored in the *i*th of m_j ordered score categories:

$$P(x_j = i | \theta_k, a_j, b_j, d_{j,1}, \ldots, d_{j,m_j-1}) = \frac{\exp(\sum_{v=0}^{i} 1.7a_j(\theta_k - b_j + d_{j,v}))}{\sum_{g=0}^{m_j-1} \exp(\sum_{v=0}^{g} 1.7a_j(\theta_k - b_j + d_{j,v}))} \qquad (11.3)$$

$$= P_{ji}(\theta_k)$$

where

m_j	is the number of categories in the response to item *j*
x_j	is the response to item *j*, with possibilities $0, 1, \ldots, m_j - 1$
a_j	is the slope parameter;
b_j	is the item location parameter characterizing overall difficulty; and
$d_{j,i}$	is the category *i* threshold parameter (see below).

Indeterminacies in the parameters of the above model are resolved by setting $d_{j,0} = 0$ and setting $\sum_{i=1}^{m_j-1} d_{j,i} = 0$. Muraki (1992) points out that $b_j - d_{j,i}$ is the point on the θ_k scale at which the plots of $P_{j,i-1}(\theta_k)$ and $P_{ji}(\theta_k)$ intersect and so characterizes the point on the θ_k scale at which the response to item *j* has equal probability of falling in response category *i*-1 and falling in response category *i*.

When $m_j = 2$, so that there are two score categories (0,1), it can be shown that $P_{ji}(\theta_k)$ of equation 11.3 for i=0,1 corresponds respectively to $P_{j0}(\theta_k)$ and $P_{j1}(\theta_k)$ of the 2PL model (equations 11.1 and 11.2 with c_j=0).

Close examination of the 3PL and generalized partial credit models indicate that both models have a linear indeterminacy of the theta scale. In other words, if the item parameters are changed, the value of θ_k can be transformed to make (11.1) and (11.3) true. For the purposes of reporting item parameter estimates and other intermediary estimates, the linear indeterminacies apparent in (11.1) and (11.3) may be resolved by an arbitrary choice of the origin and unit size in a given scale. In most cases, a provisional scale standardizing the theta distribution to have mean 0 and standard deviation 1 is employed. Final results for each content area were linearly transformed from the θ scale to a 0-to-500 scale, as described in the subject area chapters in this report.

A basic assumption of item response theory is the conditional independence of the responses by an individual to a set of items, given the individual's proficiency. That is, conditional on the individual's θ_k, the joint probability of a particular response pattern $\underline{x} = (x_1,...,x_n)$ across a set of n items is simply the product of terms based on (11.1), (11.2), and (11.3):

$$P(\underline{x}|\theta_k, item\ parameters) = \prod_{j=1}^{n} \prod_{i=0}^{m_j-1} P_{ji}(\theta_k)^{u_{ji}} \tag{11.4}$$

where $P_{ji}(\theta_k)$ is of the form appropriate to the type of item (dichotomous or polytomous), m_j is equal to 2 for the dichotomously scored items, and u_{ji} is an indicator variable defined by

$$u_{ji} = \begin{cases} 1 \text{ if response } x_j \text{ was in category } i \\ 0 \text{ otherwise.} \end{cases}$$

It is also typically assumed that response probabilities are conditionally independent of background variables ($\underline{y}$), given θ_k, or

$$P(\underline{x}|\theta_k, item\ parameters, \underline{y}) = p(\underline{x}|\theta_k, item\ parameters). \tag{11.5}$$

After $\underline{x}$ has been observed, equation 11.4 can be viewed as a likelihood function, and provides a basis for inference about θ_k or about item parameters. Estimates of item parameters were obtained by the NAEP BILOG/PARSCALE program, which combines Mislevy and Bock's (1982) BILOG and Muraki and Bock's (1991) PARSCALE computer programs, and which concurrently estimates parameters for all items (dichotomous and polytomous). The item parameters are then treated as known in subsequent calculations. In subject areas with multiple scales (reading, U.S. history, and geography), the parameters of the items constituting each of the separate scales were estimated independently of the parameters of the other scales. Once items have been calibrated in this manner, a likelihood function for the proficiency θ_k is induced by a vector of responses to any subset of calibrated items, thus allowing θ_k-based inferences from matrix samples.

In almost all NAEP IRT analyses, missing responses at the end of each block of items a student was administered were considered "not-reached," and treated as if they had not been presented to the respondent. Missing responses to dichotomous items before the last observed response in a block were considered intentional omissions, and treated as fractionally correct at the value of the reciprocal of the number of response alternatives, if the item was a multiple-choice item. These conventions are discussed by Mislevy and Wu (1988). With regard to the handling of not-reached items, Mislevy and Wu found that ignoring not-reached items introduces slight biases into item parameter estimation when not-reached items are present and speed is correlated with ability. With regard to omissions, they found that the method described above provides consistent limited-information likelihood estimates of item and ability parameters under the assumption that respondents omit only if they can do no better than responding randomly.

Occasionally, extended constructed-response items were the last item in a block of items. Because considerably more effort was required of the student to answer these items, nonresponse to an extended constructed-response item at the end of a block was considered an intentional omission (and scored as the lowest category, 0) unless the student also did not respond to the item immediately preceding that item. In that case, the extended constructed-response item was considered not reached and treated as if it had not been presented to the student.

Although the IRT models are employed in NAEP only to summarize performance, a number of checks are made to detect serious violations of the assumptions underlying the models. Checks are made to detect multidimensionality of the construct being measured and certain condition dependencies. DIF analyses are used to examine issues of dimensionality, and what are called X^2 statistics in the IRT literature are used to flag responses with serious departures from the IRT model. The latter statistics might better be called item fit statistics since they do not really have X^2 distributions. These checks include comparisons of empirical and theoretical item response functions to identify items for which the IRT model may provide a poor fit to the data. When warranted, remedial efforts are made to mitigate the effects of such violations on inferences.

Scaling areas in NAEP are determined *a priori* by grouping items into content areas for which overall performance is deemed to be of interest, as defined by the frameworks developed by the National Assessment Governing Board (NAGB). A proficiency scale θ_k is defined *a priori* by the collection of items representing that scale. What is important, therefore, is that the models capture salient information in the response data to effectively summarize the overall performance on the content area of the populations and subpopulations being assessed in the content areas. NAEP has routinely conducted differential item functioning (DIF) analyses to guard against potential biases in making subpopulation comparisons based on the proficiency distributions.

The local independence assumption embodied in equation 11.4 implies that item response probabilities depend only on θ and the specified item parameters, and not on the position of the item in the booklet, the content of items around an item of interest, or the test-administration and timing conditions. However, these effects are certainly present in any application. The practical question is whether inferences concerning aggregate performance in the scaling area that are based on the IRT probabilities obtained via 11.4 are robust with respect to the ideal assumptions underlying the IRT model. Our experience with the 1986 NAEP reading anomaly (Beaton & Zwick, 1990) has shown that for measuring small changes over time, changes in item context and speededness conditions can lead to unacceptably large random error components. These can be avoided by presenting items used to measure change in identical test forms, with identical timings and administration conditions. Thus, we do *not* maintain that the item parameter estimates obtained in any particular booklet configuration are appropriate for other conceivable configurations. Rather, we assume that the parameter estimates are context-bound. (For this reason, we prefer common population equating to common item equating whenever equivalent random samples are available for linking.) This is the reason that the data from the Trial State Assessment were calibrated separately from the data from the national NAEP—since the administration procedures differed somewhat between the Trial State Assessment and the national NAEP, the values of the item parameters could be different.

In practice, PARSCALE item fit statistics are used as a way to identify items that need further examination. Most of the statistics of this type that are available for use in this setting have distributions that are unknown. Therefore, they cannot be used for final decisions about the fit of the items to the IRT model. Because of the lack of statistical tests for IRT model fit, the fit of the IRT models to the observed data was examined within each scale by comparing the empirical item response functions (IRFs) with the theoretical curves. The primary means of accomplishing this is to generate plots of empirical versus theoretical item response curves. The theoretical curves are plots of the response functions based on the estimates of the item parameters. The empirical proportions are calculated from the posterior distributions of the thetas for each student who received the item. The plotted values are sums of these individual posteriors at each point on the theta scale for those who got the item correct plus a fraction of the omitters divided by the sum of the posteriors of those administered the item, in the case of dichotomous items, and for those who scored in the category of interest over the sum for those who received the item, in the case of polytomous items.

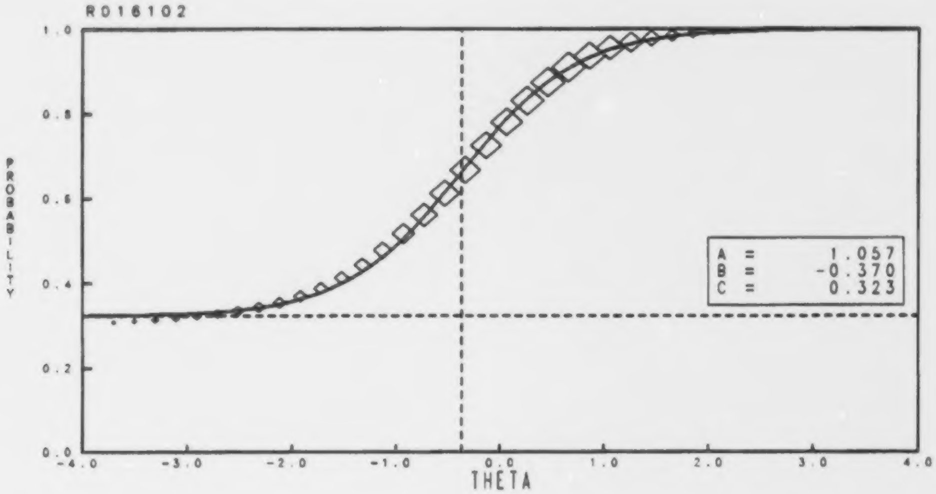
Figure 11-1 contains a plot of the empirical and theoretical IRFs for a dichotomous item. In the plot, the horizontal axis represents the theta (proficiency) scale, the vertical axis represents the probability of a correct response. The solid curve is the theoretical IRF based on the item parameter estimates and equation 11-1. The diamonds represent the empirical proportions correct as described above. The size of the diamonds are proportional to the sum of the posteriors for all of those who received the item; this is related to the number of students contributing to the estimation of that empirical proportion correct.

Figure 11-2 contains a plot of the empirical and theoretical IRFs for a polytomous item. As for the dichotomous item plot in Figure 11-1, the horizontal axis represents the proficiency scale, but the vertical axis represents the probability of having a response fall in each category. The solid curves are the theoretical IRFs based on the item parameter estimates and equation 11-3. The diamonds represent the empirical proportions of students with responses in each category and are proportional to the sum of the posteriors for the students who received the item. The heavy dashed curve represents the statistical information available from the items. This curve is what is usually called the item information curve.

For good fitting items, the empirical and theoretical curves are close together. Therefore, items for which this is not true are examined carefully. Examples of plots for specific items are provided in the subject-area chapters.

Figure 11-1

*Example Cross-Sectional Dichotomous Item (R016102, Age 13/Grade 8) Exhibiting Good Model Fit**

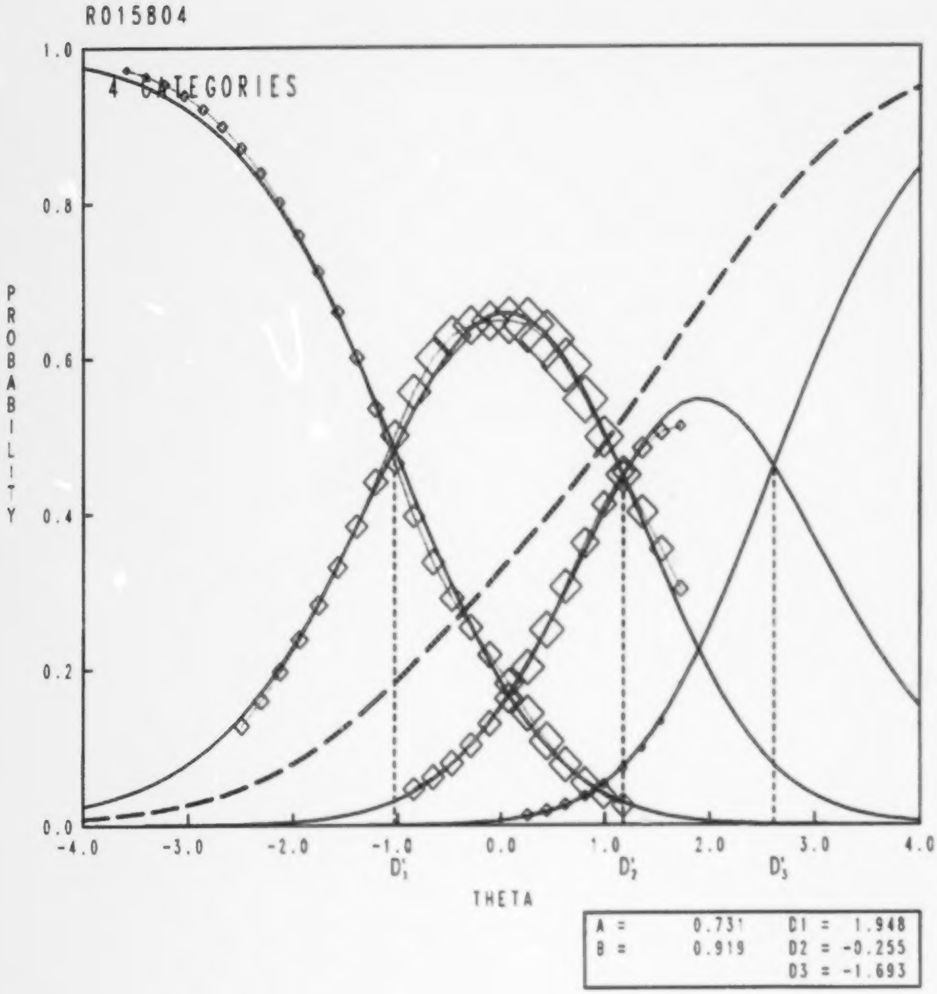


****Note:*** *The plot compares empirical and model-based estimates of the item response function (IRF). The smooth curve represents the model-based estimate at each provisional proficiency level. The diamonds represent 1994 age 13/grade 8 data.*

BEST COPY AVAILABLE

Figure 11-2

Example Cross-Sectional Polytomous Item (R015804, Age 9/Grade 4)
*Exhibiting Good Model Fit**



**__Note__: The plot compares empirical and model-based estimates of the item category response functions (ICRFs). The smooth curve represents the model-based estimate at each provisional proficiency level. The diamonds represent 1994 age 9/grade 4 data.*

11.3.2 An Overview of Plausible Values Methodology

Item response theory was developed in the context of measuring individual examinees' abilities. In that setting, each individual is administered enough items (often 60 or more) to permit precise estimation of his or her θ, as a maximum likelihood estimate $\hat{\theta}$, for example. Because the uncertainty associated with each θ is negligible, the distribution of θ, or the joint distribution of θ with other variables, can then be approximated using individuals' $\hat{\theta}$ values as if they were θ values.

This approach breaks down in the assessment setting when, in order to provide broader content coverage in limited testing time, each respondent is administered relatively few items in a subject area scale. A first problem is that the uncertainty associated with individual θs is too large to ignore, and the features of the $\hat{\theta}$ distribution can be seriously biased as estimates of the θ distribution. (The failure of this approach was verified in early analyses of the 1984 NAEP reading survey; see Wingersky, Kaplan, & Beaton, 1987.) A second problem, occurring even with test lengths of 60, arises when test forms vary across and within assessments as to the numbers, formats, and content of the test items. The measurement error distributions thus entailed differ even if underlying θ distributions do not, causing $\hat{\theta}$ distributions to exhibit spurious changes and comparisons in apparent population distributions—easily greater than actual differences over time or across groups. Although this latter problem is avoided in traditional standardized testing by presenting students with parallel test forms, controlled tightly across time and groups, the same constraints cannot be imposed in the design and data-collection phases of the present NAEP. Plausible values were developed as a way to estimate key population features consistently, and approximate others no worse than standard IRT procedures would, even when item booklet composition, format, and content balances change over time. A detailed development of plausible values methodology is given in Mislevy (1991). Along with theoretical justifications, that paper presents comparisons with standard procedures, discussions of biases that arise in some secondary analyses, and numerical examples.

The following provides a brief overview of the plausible values approach, focusing on its implementation in the 1994 NAEP analyses.

Let $\underline{y}$ represent the responses of all sampled examinees to background and attitude questions, along with variables based on the sampling design such as the school where the student is enrolled, and let $\underline{\theta}$ represent the vector of proficiency values. If $\underline{\theta}$ were known for all sampled examinees, it would be possible to compute a statistic $t(\underline{\theta},\underline{y})$—such as a scale or composite subpopulation sample mean, a sample percentile point, or a sample regression coefficient—to estimate a corresponding population quantity T. A function $U(\underline{\theta},\underline{y})$—e.g., a jackknife estimate—would be used to gauge sampling uncertainty, as the variance of t around T in repeated samples from the population.

Because the scaling models are latent variable models, however, $\underline{\theta}$ values are not observed even for sampled students. To overcome this problem, we follow Rubin (1987) by considering $\underline{\theta}$ as "missing data" and approximate $t(\underline{\theta},\underline{y})$ by its expectation given $(\underline{x},\underline{y})$, the data that actually were observed, as follows:

$$\begin{aligned} t^{*}(\underline{x},\underline{y}) &= E[t(\underline{\theta},\underline{y})|\underline{x},\underline{y}] \\ &= \int t(\underline{\theta},\underline{y})\, p(\underline{\theta}\,|\underline{x},\underline{y})\, d\underline{\theta}\,. \end{aligned} \tag{11.6}$$

It is possible to approximate t^* using random draws from the conditional distribution of the scale proficiencies given the item responses x_i, background variables y_i, and model parameters for sampled student i. These values are referred to as imputations in the sampling literature, and plausible values in NAEP. The value of $\underline{\theta}$ for any respondent that would enter into the computation of t is thus replaced by a randomly selected value from the respondent's conditional distribution. Rubin (1987) proposes that this process be carried out several times—multiple imputations—so that the uncertainty associated with imputation can be quantified. The average of the results of, for example, M estimates of t, each computed from a different set of plausible values, is a Monte Carlo approximation of (11.6); the variance among them, B, reflects uncertainty due to not observing θ, and must be added to the estimated expectation of $U(\underline{\theta},\underline{y})$, which reflects uncertainty due to testing only a sample of students from the population. Section 11.5 explains how plausible values are used in subsequent analyses.

It cannot be emphasized too strongly that **plausible values are *not* test scores for *individuals*** in the usual sense. Plausible values are offered only as intermediary computations for calculating integrals of the form of equation 11.6, in order to estimate *population* characteristics. When the underlying model is correctly specified, plausible values will provide consistent estimates of population characteristics, even though they are not generally unbiased estimates of the proficiencies of the individuals with whom they are associated. The key idea lies in the contrast between plausible values and the more familiar estimates of proficiency (e.g., maximum likelihood estimate or Bayes estimate) that are in some sense optimal for each examinee: *Point estimates that are optimal for individual examinees have distributions that can produce decidedly nonoptimal (specifically, inconsistent) estimates of population characteristics* (Little & Rubin, 1983). Plausible values, on the other hand, are constructed explicitly to provide consistent estimates of population effects. For further discussion see Mislevy, Beaton, Kaplan, and Sheehan (1992).

11.3.3 Computing Plausible Values in IRT-Based Scales

Plausible values for each respondent r are drawn from the conditional distribution $p(\underline{\theta}_r|x_r,y_r,\Gamma,\Sigma)$, where Γ and Σ are regression model parameters defined in this subsection. This subsection describes how, in IRT-based scales, these conditional distributions are characterized, and how the draws are taken. An application of Bayes' theorem with the IRT assumption of conditional independence produces

$$p(\underline{\theta}_r|x_r,y_r,\Gamma,\Sigma) \propto P(x_r|\underline{\theta}_r,y_r,\Gamma,\Sigma)\, p(\underline{\theta}_r|y_r,\Gamma,\Sigma) = P(x_r|\underline{\theta}_r)\, p(\underline{\theta}_r|y_r,\Gamma,\Sigma)\,, \tag{11.7}$$

where, for vector-valued $\underline{\theta}_r$, $P(x_r|\underline{\theta}_r)$ is the product over scales of the *independent likelihoods* induced by responses to items within each scale, and $p(\underline{\theta}_r|y_r,\Gamma,\Sigma)$ is the multivariate—and generally nonindependent—*joint density* of proficiencies for the scales, conditional on the observed value y_r of background responses, and the parameters Γ and Σ. The provisional scales are determined by the item parameter estimates that constrain the population mean to zero and standard deviation to one. The item parameter estimates are fixed and regarded as population values in the computation described in this subsection.

In the analyses of the data from the main assessments, a normal (Gaussian) form was assumed for $p(\underline{\theta}|y_i,\Gamma,\Sigma)$, with a common variance-covariance matrix, Σ, and with a mean given by a linear model with slope parameters, Γ, based on the first approximately 200 principal components of several hundred selected main-effects and two-way interactions of the complete vector of background variables. The included principal components will be referred to as the *conditioning variables*, and will be denoted y^c.

(The complete set of original background variables used in the analyses of each subject area are listed in Appendix C.) The following model was fit to the data within each subject area:

$$\underline{\theta} = \Gamma \underline{y} + \underline{\varepsilon}, \tag{11.8}$$

where $\underline{\varepsilon}$ is multivariately normally distributed with mean zero and variance-covariance matrix Σ. The number of principal components of the background variables used for each sample was sufficient to account for 90 percent of the total variance of the full set of background variables (after standardizing each variable). As in regression analysis, Γ is a matrix each of whose columns is the *effects* for one scale and Σ is the matrix *variance-covariance of residuals* between scales.

A model similar to (11.8) was used for the long-term trend assessments, with the difference that $\underline{y}$ consisted of main effects and interactions from the smaller set of background variables (rather than principal components of those variables) available in the long-term trend assessments.

Maximum likelihood estimates of Γ and Σ, denoted by $\hat{\Gamma}$ and $\hat{\Sigma}$, are obtained with extensions of Sheehan's (1985) MGROUP computer program using the EM algorithm described in Mislevy (1985). The EM algorithm requires the computation of the mean, $\bar{\underline{\theta}}_r$, and variance-covariance matrix, Σ_r^p, of the posterior distribution in (11.7) for respondent *i* when there are *p* scales within a subject area. For subject areas with multiple scales, the CGROUP version of the MGROUP program was used to compute the moments using higher order asymptotic corrections to a normal approximation (Thomas, 1993). For the long-term trend assessments, each of which have a single scale, the more precise but computationally intensive BGROUP version of MGROUP was used. BGROUP uses numeric quadrature to evaluate the posterior moments required by the E-step of the EM algorithm for one- and two-dimensional applications (Thomas, 1993). For estimation of group means on a single scale, CGROUP and BGROUP results will be nearly identical to those from the original MGROUP program. CGROUP and BGROUP yield better estimates of correlations between scales, and hence better estimates of composite scale means. BGROUP will, theoretically, yield better estimates than CGROUP, but because of the heavy computational demands of the methodology used, its function is limited to bivariate scales. Hence CGROUP is used for assessments involving more than two scales.

After completion of the EM algorithm, the plausible values are drawn in a three-step process from the joint distribution of the values of Γ for all sampled respondents. First, a value of Γ is drawn from a normal approximation to $P(\Gamma,\Sigma | x_r, y_r)$ that fixes Σ at the value $\hat{\Sigma}$, (Thomas, 1993). Second, conditional on the generated value of Γ (and the fixed value of $\Sigma = \hat{\Sigma}$), the mean, $\bar{\underline{\theta}}_r$, and variance-covariance matrix, Σ_r^p, of the posterior distribution (of respondent *i* when there are *p* scales within the subject area) in equation 11.7 (i.e., $p(\underline{\theta}_r | x_r, y_r, \Gamma, \Sigma)$) are computed using the same methods applied in the EM algorithm. In the third step, the $\underline{\theta}_r$ are drawn independently from a multivariate normal distribution with mean $\bar{\underline{\theta}}_r$ and variance-covariance matrix Σ_r^p, approximating the distribution in (11.7). These three steps are repeated five times producing five imputations of $\bar{\underline{\theta}}_r$ for each sampled respondent.

11.4 ANALYSES

When survey variables are observed without error from every respondent, standard variance estimators quantify the uncertainty associated with sample statistics from the only source of uncertainty, namely the sampling of respondents. Item-level statistics for NAEP cognitive items meet this requirement, but proficiency values do not. The IRT models used in their construction posit an unobservable proficiency variable θ to summarize performance on the items in the subarea. The fact that θ values are not observed even for the respondents in the sample requires additional statistical analyses to draw inferences about θ distributions and to quantify the uncertainty associated with those inferences. As described above, Rubin's (1987) multiple imputations procedures were adapted to the context of latent variable models to produce the plausible values upon which many analyses of the data from the 1994 assessment were based. This section describes how plausible values were employed in subsequent analyses to yield inferences about population and subpopulation distributions of proficiencies.

11.4.1 Computational Procedures

Even though one does not observe the θ value of respondent r, one does observe variables that are related to it: x_r, the respondent's answers to the cognitive items he or she was administered in the area of interest, and y_r, the respondent's answers to demographic and background variables. Suppose one wishes to draw inferences about a number $T(\underline{\theta},\underline{Y})$ that could be calculated explicitly if the θ and y values of each member of the population were known. Suppose further that if θ values were observable, we would be able to estimate T from a sample of N pairs of θ and y values by the statistic $t(\underline{\theta},\underline{y})$ [where $(\underline{\theta},\underline{y}) \equiv (\theta_1,y_1,\ldots,\theta_N,y_N)$], and that we could estimate the variance in t around T due to sampling respondents by the function $U(\underline{\theta},\underline{y})$. Given that observations consist of (x_i,y_i) rather than $(\underline{\theta}_i,y_i)$, we can approximate t by its expected value conditional on $(\underline{x},\underline{y})$, or

$$t^*(\underline{x},\underline{y}) = E[t(\underline{\theta},\underline{y})|\underline{x},\underline{y}] = \int t(\underline{\theta},\underline{y})\, p(\underline{\theta}|\underline{x},\underline{y})\, d\underline{\theta}\,.$$

It is possible to approximate t^* with random draws from the conditional distributions $p(\underline{\theta}_i|x_i,y_i)$, which are obtained for all respondents by the method described in Section 11.3.3. Let $\hat{\underline{\theta}}_m$ be the mth such vector of plausible values, consisting of a multidimensional value for the latent variable of each respondent. This vector is a plausible representation of what the true $\underline{\theta}$ vector might have been, had we been able to observe it.

The following steps describe how an estimate of a scalar statistic $t(\underline{\theta},\underline{y})$ and its sampling variance can be obtained from M (>1) such sets of plausible values. (Five sets of plausible values are used in NAEP analyses.)

1) Using each set of plausible values $\hat{\underline{\theta}}_m$ in turn, evaluate t as if the plausible values were true values of $\underline{\theta}$. Denote the results $\hat{t}_m$, for $m=1,\ldots,M$.

2) Using the jackknife variance estimator defined in Chapter 10, compute the estimated sampling variance of $\hat{t}_m$, denoting the result U_m.

3) The final estimate of t is

$$t^* = \sum_{m=1}^{M} \frac{\hat{t}_m}{M}.$$

4) Compute the average sampling variance over the M sets of plausible values, to approximate uncertainty due to sampling respondents:

$$U^* = \sum_{m=1}^{M} \frac{U_m}{M}.$$

5) Compute the variance among the M estimates $\hat{t}_m$, to approximate uncertainty due to not observing θ values from respondents:

$$B = \sum_{m=1}^{M} \frac{(\hat{t}_m - t^*)^2}{(M - 1)}$$

6) The final estimate of the variance of t^* is the sum of two components:

$$V = U^* + (1 + M^{-1})\, B.$$

Note: Due to the excessive computation that would be required, NAEP analyses did not compute and average jackknife variances over all five sets of plausible values, but only on the first set. Thus, in NAEP reports, U^* is approximated by U_1.

11.4.2 Statistical Tests

Suppose that if θ values were observed for sampled students, the statistic $(t - T)/U^{1/2}$ would follow a t-distribution with d degrees of freedom. Then the incomplete-data statistic $(t^* - T)/V^{1/2}$ is approximately t-distributed, with degrees of freedom given by

$$\nu = \frac{1}{\dfrac{f^2}{M - 1} + \dfrac{(1 - f)^2}{d}}$$

where f is the proportion of total variance due to not observing θ values:

$$f = (1+M^{-1})\, B/V.$$

When B is small relative to U^*, the reference distribution for incomplete-data statistics differs little from the reference distribution for the corresponding complete-data statistics. This is the case with main NAEP reporting variables. If, in addition, d is large, the normal approximation can be used to flag "significant" results.

For k-dimensional t, such as the k coefficients in a multiple regression analysis, each U_m and U^* is a covariance matrix, and B is an average of squares and cross-products rather than simply an average of squares. In this case, the quantity $(T-t^*)\ V^{-1}\ (T-t^*)'$ is approximately F distributed, with degrees of freedom equal to k and ν, with ν defined as above but with a matrix generalization of f:

$$f = (1+M^{-1})\ Trace\ (BV^{-1})/k\ .$$

By the same reasoning as used for the normal approximation for scalar t, a chi-square distribution on k degrees of freedom often suffices for multivariate t.

For practical purposes, in production of NAEP comparisons, the impact of the reduced degrees of freedom on inferential techniques can be largely accounted for by (1) using a moderate number (say 30) of degrees of freedom for all inferences about subgroups that appear approximately uniformly in all PSUs, and (2) using a smaller number (say 10) of the remaining subgroups. Certainly one should be cautious about barely significant results for subgroups that are highly clustered in the population.

11.4.3 Biases in Secondary Analyses

Statistics t^* that involve proficiencies in a scaled content area and variables included in the conditioning variables y^c are consistent estimates of the corresponding population values T. This includes interrelationships among scales within a content area that have been treated in the multivariate manner described above in Section 11.3.3. Statistics involving background variables y that were *not* conditioned on, or relationships among proficiencies from *different* content areas, are subject to asymptotic biases whose magnitudes depend on the type of statistic and the strength of the relationships of the nonconditioned background variables to the variables that were conditioned on and to the proficiency of interest. That is, the large sample expectations of certain sample statistics need not equal the true population parameters.

The *direction* of the bias is typically to underestimate the effect of nonconditioned variables. For details and derivations see Beaton and Johnson (1990), Mislevy (1991), and Mislevy and Sheehan (1987, Section 10.3.5). For a given statistic t^* involving one content area and one or more nonconditioned background variables, the *magnitude* of the bias is related to the extent to which observed responses x account for the latent variable θ, and the degree to which the nonconditioned background variables are explained by conditioning background variables. The first factor—conceptually related to test reliability—acts consistently in that greater measurement precision reduces biases in *all* secondary analyses. The second factor acts to reduce biases in certain analyses but increase it in others. In particular,

- High shared variance between conditioned and nonconditioned background variables *mitigates* biases in analyses that involve only proficiency and nonconditioned variables, such as marginal means or regressions.

- High shared variance *exacerbates* biases in regression coefficients of conditional effects for nonconditioned variables, when nonconditioned and conditioned background variables are analyzed jointly as in multiple regression.

The large number of background variables that have been included in the conditioning vectors for the 1994 assessments allows a large number of secondary analyses to be carried out with little or no bias, and mitigates biases in analyses of the marginal distributions of θ in nonconditioned variables. Kaplan and Nelson's analysis of the 1988 NAEP reading data (some results of which are summarized in Mislevy, 1991), which had a similar design and fewer conditioning variables, indicates that the potential bias for nonconditioned variables in multiple regression analyses is below 10 percent, and biases in simple regression of such variables is below five percent. Additional research (summarized in Mislevy, 1990) indicates that most of the bias reduction obtainable from conditioning on a large number of variables can be captured by instead conditioning on the first several principal components of the matrix of all original conditioning variables. This procedure was adopted for the 1992 and 1994 main assessments by replacing the conditioning effects by the first *K* principal components, where *K* was selected so that 90 percent of the total variance of the full set of conditioning variables (after standardization) was captured. Mislevy (1990) shows that this puts an upper bound of 10 percent on the average bias for all analyses involving the original conditioning variables.

11.4.4 A Numerical Example

To illustrate how plausible values are used in subsequent analyses, this subsection gives some of the steps in the calculation of the 1992 grade 4 reading composite mean and its estimation-error variance.

The weighted mean of the first plausible values of the reading composite for the grade 4 students in the sample is 217.79, and the jackknife variance of these values is 0.833. Were these values true θ values, then 217.79 would be the estimate of the mean and 0.833 would be the estimation-error variance. The weighted mean of the second plausible values of the same students, however, is 217.62; the third, fourth, and fifth plausible values give weighted means of 217.74, 218.24, and 218.05. Since all of these figures are based on precisely the same sample of students, the variation among them is due to uncertainty about the students' θs, having observed their item responses and background variables. Consequently, our best estimate of the mean for grade 4 students is the average of the five plausible values: 217.89. Taking the jackknife variance estimate from the first plausible value, 0.833, as our estimate U^* of sampling variance, and the variance among the five weighted means, .063, as our estimate B of uncertainty due to not observing θ, we obtain as the final estimate V of total error variance $0.833 + (1+5^{-1})\ .063 = 0.908$.

It is also possible to partition the estimation error variance of a statistic using these same variance components. The proportion of error variance due to sampling students from the population is U^*/V, and the proportion due to the latent nature of θ is $(1+M^{-1})B/V$. The results are shown in Table 11-1. The value of U^*/V roughly corresponds to reliability in classical test theory and indicates the amount of information about an average individual's θ present in the observed responses of the individual. It should be recalled again that the objective of NAEP is not to estimate and compare values of individual examinees, the accuracy of which is gauged by reliability coefficients. The objective of NAEP, rather, is to estimate population and subpopulation characteristics, and the marginal estimation methods described above have been designed to do so consistently regardless of the values of reliability coefficients.

Table 11-1

Estimation Error Variance and Related Coefficients for the 1992 Grade 4 Reading Composite (Based on Five Plausible Values)

			Proportion of Variance Due to...	
U*	$(1+5^{-1})B$	V	Student Sampling: U*/V	Latency of θ: $(1+5^{-1})B/V$
0.833	0.076	0.908	0.92	0.08

Chapters 12 through 18 and Appendix E provide values of the proportion of variance due to sampling and due to the latent nature of θ for all 1994 scales and composites for the populations as a whole and, in the appendix, for selected subpopulations. It will be seen that the proportion of variance due to the latency of θ varies somewhat among subject areas, tending to be largest for the long-term trend writing assessment, where there is low correlation between tasks and each student responded to only one or at most two tasks. The proportion of variance due to latency of θ is smallest for the composites of the main assessment subjects, where the number of items per student is largest. Essentially, the variance due to the latent nature of θ is largest when there is less information about a student's proficiency. (Note the distinction between estimation error variance of a parameter estimate and the estimate of the variance of the θ distribution. The former depends on the accuracy of measurement; the large-sample model-based expected value of the latter does not.) Given fixed assessment time, this decrease in information will occur whenever the amount of information per unit time decreases as can happen when many short constructed-response or multiple-choice items are replaced by a few extended constructed-response items.

11.5 DESCRIBING STUDENT PERFORMANCE

Since its beginning, a goal of NAEP has been to inform the public about what students in American schools know and can do. While the NAEP scales provide information about the distributions of proficiency for the various subpopulations, they do not directly provide information about the meaning of various points on the scale. Traditionally, meaning has been attached to educational scales by norm-referencing—that is, by comparing students at a particular scale level to other students. In contrast, NAEP achievement levels and scale anchors describe selected points on the scale in terms of the types of skills that are likely to be exhibited by students scoring at that level. Both the achievement level process and scale anchoring of certain percentiles of the student proficiency distribution were applied to the U.S. history and geography composites. Scale anchoring of certain percentiles of the student proficiency distribution was applied to the reading composite using the 1994 data. However, the achievement level process was completed for the 1992 reading assessment, so the results were directly applied to the 1994 results. In addition, each item was mapped to a point on the scale in which it belonged, so that the content of each item provides information about what students at each score level can do in a probabistic sense.

11.5.1 Achievement Levels

NAGB has determined that achievement levels shall be the first and primary way of reporting NAEP results. Setting achievement levels is a method for setting standards on the NAEP assessment that identify what students *should* know and be able to do at various points on the composite. For each grade of each subject, three levels were defined—basic, proficient, and advanced. Based on initial policy definitions of these levels, panelists were asked to determine operational descriptions of the levels appropriate with the content and skills assessed in the assessment. With these descriptions in mind, the panelists were then asked to rate the assessment items in terms of the expected performance of marginally acceptable examinees at each of these three levels. These ratings were then mapped onto the NAEP scale to obtain the achievement level cutpoints for reporting. Further details of the achievement level-setting process for U.S. history and geography appear in Appendix G.

11.5.2 Performance Descriptions Based on Composite Scales

A procedure known as scale anchoring was used to develop descriptions of student performance at selected points on the NAEP reading, U.S. history, and geography composite scales. The scale points that were selected for anchoring reflect three levels of knowledge and abilities corresponding to lower-, middle-, and higher-performing students for each subject. In U.S. history and geography, these levels correspond to the 25th, 50th, and 90th percentile points on the composite scale as established by the performance of students in 1994. For reading, the selected points correspond to the 25th, 50th, and 90th percentile points on the composite scale as established by the performance of students in 1992 and reported in the *NAEP 1992 Reading Report Card for the Nation and the States, Data from the National and Trial State Assessments* (Mullis, Campbell, & Farstrup, 1993).

Around each percentile point, a band was built to define a range of scale scores. Students described as being at a particular level were within a five percentile point range on either side of the specified scale point. For example, the 50th percentile was defined as the region between the 45th and 55th percentile points on the scale. A question was identified as anchoring at a percentile point on the scale if it was answered successfully by at least 65 percent of the students within that percentile band. (The criterion was set at 74 percent for multiple-choice questions to correct for the possibility of answering correctly by guessing.)

After defining the bands of the scale to be anchored, the next step in the process was to identify: (1) questions answered correctly for dichotomously scored questions, or (2) questions answered at a particular score level for partial credit constructed-response questions. Because the extended constructed-response questions were scored according to four levels of performance, each extended constructed-response question was treated as three distinct questions corresponding to scores of Partial or better, Essential or better, and Extensive. These distinct score levels were then analyzed in the same manner as questions scored dichotomously, as either correct or incorrect. Thus, for example, an extended constructed-response question might anchor at the 50th percentile for Partial or better responses **and** at the 90th percentile for Essential or better responses.

A committee of subject area experts, including teachers for the grades involved, college professors, state curriculum supervisors, and researchers, was assembled to review the sets of questions identified for each percentile band. The committee was divided into three groups, one for each grade. Each group examined and analyzed questions that anchored at the 25th, 50th, and 90th percentiles to determine the specific knowledge and abilities associated with each question.

Committee members were also provided with the sets of questions at each grade that "did not anchor" to inform their decisions about what students could do by seeing examples of what they could not do. Drawing on their knowledge of the subject area, committee members were asked to summarize student performance by describing the knowledge, skills, and abilities demonstrated by students in each of the score bands.

The performance descriptions are cumulative; that is, the abilities described for the lower performing students are considered to be among the abilities of students performing at higher points on the scale. Therefore, the full description of students' knowledge and abilities in the middle scale band would include those abilities described at the lower band. Similarly, the abilities of students performing at the higher scale band include the abilities described for students at the middle and lower bands.

11.5.3 Item Mapping Procedures

In order to map items (questions) to particular points on each subject area scale, a response probability convention had to be adopted that would divide those who had a higher probability of success from those who had a lower probability. Establishing a response probability convention has an impact on the mapping of assessment items onto the scales. A lower boundary convention maps the items at lower points along the scales, and a higher boundary convention maps the same items at higher points along the scales. The underlying distribution of skills in the population does not change, but the choice of a response probability convention does have an impact on the proportion of the student population that is reported as "able to do" the items on the scales.

There is no obvious choice of a point along the probability scale that is clearly superior to any other point. If the convention were set with a boundary at 50 percent, those above the boundary would be more likely to get an item right than get it wrong, while those below that boundary would be more likely to get the item wrong than right. While this convention has some intuitive appeal, it was rejected on the grounds that having a 50/50 chance of getting the item right shows an insufficient degree of mastery. If the convention were set with a boundary at 80 percent, students above the criterion would have a high probability of success with an item. However, many of the students below this criterion show some level of achievement that would be ignored by such a stringent criterion. In particular, those in the range between 50 and 80 percent correct would be more likely to get the item right than wrong, yet would not be in the group described as "able to do" the item.

In a compromise between the 50 percent and the 80 percent conventions, NAEP has adopted two related response probability conventions: 74 percent for multiple-choice items (to correct for the possibility of answering correctly by guessing), and 65 percent for constructed-response items (where guessing is not a factor). These probability conventions were established, in part, based on an intuitive judgment that they would provide the best picture of students' knowledge and skills.

Some additional support for the dual conventions adopted by NAEP was provided by Huynh (1994). He examined the IRT information provided by items, according to the IRT model used in scaling NAEP items. Following Bock (1972), Huynh decomposed the item information into that provided by a correct response [$P(\theta) * I(\theta)$] and that provided by an incorrect response [$(1-P(\theta)) * I(\theta)$]. Huynh showed that the item information provided by a correct response to a constructed-response item is maximized at the point along the scale at which two-thirds of the students get the item correct (for multiple-choice items, information is maximized at the point at which 74 percent get the item correct). *It should be noted, however, that maximizing the item information $I(\theta)$, rather than the information provided by a correct response $[P(\theta) * I(\theta)]$, would imply an item mapping criterion closer to 50 percent.*

For dichotomously-scored items the information function as defined by Birnbaum (1968, p. 463) is defined for the *j*th item as

$$I_j(\theta) = \frac{(1.7a_j)^2 P_{j0}(\theta)[P_{j1}(\theta) - c_j]^2}{P_{j1}(\theta)(1 - c_j)^2}$$

where the notation is the same as that used in equations 11.1 and 11.2. The item information function was defined by Samejima (1969) in general for polytomously-scored items, and has been derived for items scaled by the generalized partial credit model (Muraki, 1993a; Muraki, 1993b) as (in a slightly different, but equivalent form)

$$I_j(\theta) = (1.7a_j)^2 \left[\sum_{i=0}^{m_j-1} i^2 P_{ji}(\theta) - \left\{ \sum_{i=0}^{m_j-1} i P_{ji}(\theta) \right\}^2 \right].$$

11.6 OVERVIEW OF THE 1994 NAEP SCALES

The following IRT scale-score analyses were carried out for the 1994 NAEP assessment:

- *Reading:* Three IRT scales linking back to the 1992 main assessment of reading and one IRT scale linking 1994 results to results from reading assessments in 1971, 1975, 1979, 1984, 1988, 1990, and 1992.
- *U.S. History:* Four newly developed IRT scales for the main assessment of U.S. history.
- *Geography:* Three newly developed IRT scales for the main assessment of geography.
- *Mathematics:* One unidimensional IRT mathematics scale linking 1994 results to results from mathematics assessments in 1973, 1976, 1982, 1986, 1990, and 1992.
- *Science:* One unidimensional scale linking 1994 results to results from science assessments in 1969, 1973, 1977, 1982, 1986, 1990, and 1992.
- *Writing:* One polytomous item scale linking 1994 writing results to the 1984, 1988, 1990, and 1992 assessments.

Details follow in Chapters 12 through 18.

Chapter 12

DATA ANALYSIS FOR THE READING ASSESSMENT[1]

John R. Donoghue, Steven P. Isham, and Lois H. Worthington
Educational Testing Service

12.1 INTRODUCTION

This chapter describes the analyses performed on the responses to the cognitive and background items in the 1994 assessment of reading. These analyses led to the results presented in the *1994 NAEP Reading: A First Look, Findings from the National Assessment of Educational Progress* (Williams, Reese, Campbell, Mazzeo, & Phillips, 1995) and the *NAEP 1994 Reading Report Card for the Nation and the States, Findings from the National Assessment of Educational Progress and Trial State Assessments* (Campbell, Donahue, Reese, & Phillips, 1996) . The emphasis of this chapter is on the methods and results of procedures used to develop the IRT-based scale scores that formed the basis of these reports. However, some attention is given to the analysis of constructed-response items as reported in the *NAEP 1994 Reading Report Card for the Nation and the States, Findings from the National Assessment of Educational Progress and Trial State Assessments.* The theoretical underpinnings of the IRT and plausible values methodology described in this chapter are given in Chapter 11, and several of the statistics are described in Chapter 9.

The objectives of the reading analyses were to:

- Prepare scale values for the analysis of the main focused-BIB reading samples.
- Link the 1994 main focused-BIB samples to the 1992 reading scale and perform all analyses necessary to produce a short-term trend report in reading. The reading short-term trend results include the years 1992 and 1994.
- Prepare the analysis of the *NAEP Reader*. The *NAEP Reader* was a special 50-minute cognitive block administered at grades 8 and 12. Each student was given a booklet of seven short stories, followed by a set of generic questions. One sample of students was given the choice of which story to read. Another sample of students was told which of the stories in the booklet to read. Each student then responded to the same generic questions. Data from the *NAEP Reader* were analyzed separately, and were *not* incorporated into the Reading for Literary Experience scale. The *NAEP Reader* sample will not be discussed further in this chapter; results will be reported in a special, focused report. The design of the *NAEP Reader* study will provide valuable information about the role of examinee choice in assessment.

---

[1] John R. Donoghue was the primary person responsible for the planning, specification, and coordination of the national reading analyses. Computer activities for all reading scaling and data analyses were directed by Steven P. Isham and performed by Steven P. Isham and Lois H. Worthington. Nancy L. Allen, John Mazzeo, and Robert Mislevy consulted on IRT scaling and generation of plausible values. Hua-hua Chang assisted with some analyses.

The major analysis components are discussed in turn. Some aspects of the analysis, such as procedures for item analysis, scoring of constructed-response items, and methods of scaling, are described in previous chapters and are, therefore, not detailed here.

The student samples that were administered reading items in the 1994 assessment are shown in Table 12-1. (See Chapters 1 and 3 for descriptions of the target populations and the sample design used for the assessment.) Data from the samples denoted Rdg-MainP were used in the main analysis. Data from the 1994 main focused-BIB samples were scaled separately from the data from the long-term trend samples that contributed to the trends in reading achievement. Accordingly, the trend and main analyses are presented in separate chapters. This chapter contains information about the scaling of the data from the main focused-BIB samples and the formation of the short-term trend; Chapter 15 pertains to the scaling of the data from the reading long-term trend.

This chapter describes the analyses carried out in the development of the 1994 main reading scales. The procedures used were similar to those employed in the analysis of the 1992 assessment of reading (Donoghue, Isham, Bowker, & Freund, 1994), and are based on the philosophical and theoretical underpinnings described in Chapter 11. The next sections contain a description of the analysis performed using the main focused-BIB sample data.

There were six major steps in the analysis of the 1994 main reading assessment, each of which is described in a separate section:

- conventional item and test analysis (Section 12.3.1);
- differential item functioning analysis (Section 12.3.2);
- item response theory (IRT) scaling (Section 12.4);
- estimation of national and subgroup proficiency distributions based on the 'plausible values' methodology (Section 12.5);
- transformation of the purposes-for-reading scales to the proficiency metric (Section 12.6.1); and
- creation of the composite reading scale (Section 12.6.2).

12.2 DESCRIPTION OF STUDENT SAMPLE, ITEMS AND ASSESSMENT BOOKLETS

The data from the main focused-BIB assessment of reading (9[Rdg-MainP], 13[Rdg-MainP], and 17[Rdg-MainP]) were used for main analyses comparing the levels of reading achievement for various subgroups of the 1994 target populations. The main assessment included three student cohorts: students who were either in the fourth-grade or 9 years old, students who were either in the eighth grade or 13 years old, and students who were either in the twelfth grade or 17 years old. The birthdate ranges for age-eligible students were based on the 1984, 1980, and 1976 calendar years respectively for ages 9, 13, and 17. The sampled students in each of these three cohorts were assessed in the winter. The samples in the main assessment are listed in Table 12-1.

Table 12-1
NAEP 1994 Reading Student Samples

Sample	Booklets	Mode	Cohort Assessed	Time of Testing	Age Definition	Modal Grade	Number Assessed
9 [Rdg-MainP]	1-16	Print	Age 9/ Grade 4	1/3/94- 4/1/94	CY	4	9,680
13 [Rdg-MainP]	1-18, 21	Print	Age 13/ Grade 8	1/3/94- 4/1/94	CY	8	13,547
13 [Rdg-MainR]	19, 20	Print	Age 13/ Grade 8	1/3/94- 4/1/94	CY	8	7,325
17 [Rdg-MainP]	1-18, 21, 22	Print	Age 17/ Grade 12	1/3/94- 4/1/94	CY	12	13,705
17 [Rdg-MainR]	19, 20	Print	Age 17/ Grade 12	1/3/94- 4/1/94	CY	12	5,815

LEGEND

Rdg	*Reading*	*Print*	*Print administration*
MainP	*Main assessment, print administration*	*CY*	*Calendar year: birthdates in 1984, 1980, and 1976, respectively for ages 9, 13, and 17.*
MainR	*NAEP Reader study*		

The items in the assessment were based on the curriculum framework described in *Reading Framework for the 1992 and 1994 National Assessment of Educational Progress* (National Assessment Governing Board, 1994). The 1994 reading assessment is based on the same objectives as the 1992 cross-sectional assessment. Compared to earlier NAEP assessments, the current assessment contains longer reading passages that are intended to be more authentic examples of the reading tasks encountered in and out of school. As described in the reading framework, these blocks are organized into three scales, corresponding to three purposes for reading: Reading for Literary Experience, Reading to Gain Information, and Reading to Perform a Task. At age 9/grade 4, only the first two purposes are represented. Scales were produced for each of the purposes of reading. In addition, a composite scale for reading was created as a weighted sum of the purposes-for-reading scales.

In the main samples, each student was administered a booklet containing two separately timed 25-minute blocks of cognitive reading items. In addition, each student was administered a block of background questions, a block of reading-related background questions, and a block of questions concerning the student's motivation and his or her perception of the difficulty of the cognitive items. The background and motivational blocks were common to all reading booklets for a particular age/grade level. Eight (age 9/grade 4) or nine (age 13/grade 8 and age 17/grade 12) 25-minute blocks of reading items were administered at each age/grade level. As described in Chapter 2, the 25-minute blocks were combined into booklets according to a partially balanced incomplete block (BIB) design. (See Chapter 4 for more information about the blocks and booklets.) In addition, 50-minute reading blocks (in lieu of the two 25-minute blocks) were presented to the older students, one at age 13/grade 8 and two at age 17/grade 12.

The 50-minute blocks were closely examined to ensure the appropriateness of including them with the shorter blocks in the scaling.[2] Both age- and grade-selected students contributed to the main scaling. However, the 'grade-only' portion of the main focused-BIB reading samples contributed to the means and percentages of the main results that are reported in the *NAEP 1994 Reading Report Card for the Nation and the States: Findings from the National Assessment of Educational Progress and Trial State Assessments.*

For each age/grade, approximately two-thirds of the items in the cross-sectional assessment were identical to items in the 1992 cross-sectional assessment. These items occurred in intact blocks, and provided the common information needed to establish the short-term trend. Table 12-2 gives the blocks and numbers of items common across assessment years.

Table 12-2

1994 Reading Cross-Sectional Blocks and Items Common to the 1992 Assessment

Sample	Common Blocks	Number of Common Items
9[Rdg-MainP]	R3, R4, R5, R6, R7, R10	66
13[Rdg-MainP]	R3, R5, R6, R7, R10, R11	71
17[Rdg-MainP]	R3, R4, R6, R7, R10, R11, R13*	87

****Note:*** *50-minute block*

Each block consisted of one or two reading passages, each followed by several items. In addition to multiple-choice items, each block contained a number of constructed-response items, accounting for well over half of the testing time. Constructed-response items were scored by specially trained readers (described in Chapter 7). Some of the constructed-response items required only a few sentences or a paragraph response. These short constructed-response items were scored dichotomously as correct or incorrect. Other constructed-response items (regular constructed-response items) required somewhat more elaborated responses, and were scored polytomously on a 3-point (0-2) scale:

0 = Unsatisfactory (and omit)
1 = Partial
2 = Complete

---

[2] These analyses were identical to those described in *Assessing Some of the Properties of Longer Blocks in the 1992 NAEP Reading Assessment* (Donoghue & Mazzeo, 1993). Additional comparisons based on bootstrap comparisons (Donoghue, 1995) further supported the comparability of the 25- and 50-minute reading blocks.

BEST COPY AVAILABLE

In addition, each block contained at least one extended constructed-response item, which required a more in-depth, elaborated response. These items were scored polytomously on a 4-point (0-3) scale:

0 = Unsatisfactory (and omit)
1 = Partial
2 = Essential
3 = Extensive, which demonstrates more in-depth understanding

Originally, the scoring guides for regular constructed-response (3-point) items and extended constructed-response (4-point) items separated the "unsatisfactory" from the "omit" responses, with omits and off-task responses forming a category below the "unsatisfactory" responses (the treatment of items that were not reached is discussed below in Section 12.3.1). During the 1992 scaling process, it was discovered that this scoring rule resulted in unexpectedly poor fit to the IRT model. After much investigation, the 0 category (omitted and off-task responses) was recoded. Off-task responses were treated as "not administered" for each of the items, and omitted responses were combined with the next lowest category, "unsatisfactory." For new items (administered for the first time in 1994), decisions concerning the treatment of omit and off-task responses were re-examined and found to be appropriate for these new items.

In addition, adjacent categories of a small number of constructed-response items were combined (collapsed). These changes were made so that the scaling model used for these items fit the data more closely, and are described more fully in Section 12.4.3.1.1. Some of the short-term trend items had been collapsed in the original 1992 scaling. These items were collapsed in an identical manner for the 1994 assessment. New items (unique to 1994) were also examined, and where necessary, adjacent categories were collapsed.

The composition of each block of items, in terms of format and content, is given in Table 12-3. Common labeling of these blocks across age/grade levels does not denote common items. The numbers of items scaled in 1994 for each age/grade are presented in Table 12-4.

12.3 ITEM ANALYSES

12.3.1 Conventional Item and Test Analyses

Tables 12-5, 12-6, and 12-7 show the number of items in the block, the average weighted item score, average weighted polyserial correlation, and the weighted alpha reliability for each block administered. These statistics are described in Chapter 9. These values were calculated for the items within each block used in the scaling process. The tables also give the number of students who were administered the block and the percent not reaching the last item in the block. These numbers include only those students in the grade-only portion of the samples that contributed to the summary statistics provided in the *NAEP 1994 Reading Report Card for the Nation and the States: Findings from the National Assessment of Educational Progress and Trial State Assessments.* Student weights were used for all statistics, except for the sample sizes. The results for the blocks administered to each age/grade level indicate that the blocks differ in number of items, average difficulty, reliability, and percent not reaching the last item, and so are not parallel to each other. Preliminary item analyses for all items within a block were completed before scaling; however, the results shown here indicate the characteristics of the items that contributed to the final scale, and reflect decisions made in scaling to combine adjacent categories (collapse) for a small number of items.

As described in Chapter 9, in NAEP analyses (both conventional and IRT-based), a distinction is made between missing responses at the end of each block (not reached) and missing responses prior to the last observed response (omitted). Items that were not reached were treated as if they had not been presented to the examinee, while omitted items were regarded as incorrect. The proportion of students attempting the last item of a block (or, equivalently, one minus the proportion not reaching the last item) is often used as an index of the degree of speededness of the block of items.

Standard practice at ETS is to treat all nonrespondents to the last item as if they had not reached the item. For multiple-choice items, short constructed-response items, and regular constructed-response items, this convention produced a reasonable pattern of results, in that the proportion reaching the last item does not differ markedly from the proportion attempting the next-to-last item. However, for the blocks that ended with extended constructed-response items, this convention resulted in an implausibly large drop in the number of students attempting the final item. Therefore, for blocks that ended with an extended constructed-response item, students who attempted the next-to-last item but did not respond to the last item were classified as having intentionally omitted that item.

The results in Tables 12-5 to 12-7 indicate that the difficulty and internal consistency of the blocks varied. Such variability is expected, because the blocks were not constructed to be parallel. Based on the proportion of students attempting the last item, all of the blocks appear to be somewhat speeded. This effect is larger for grade 4 than for the other grades.

Table 12-3

Item Type Composition of Blocks Used in the Reading Main Assessment

Grade	Type of Item	Block										
		R3	R4	R5	R6	R7	R8	R9	R10	R11	R13	R14
4	Number of multiple-choice items	6	5	7	5	4	3	3	6	—	—	—
	Number of short constructed-response items	4	6	3	4	5	0	1	5	—	—	—
	Number of regular constructed-response items	0	0	0	0	0	5	4	0	—	—	—
	Number of extended constructed-response items	1	1	1	1	1	1	1	1	—	—	—
	Total	11	12	11	10	10	9	9	12	—	—	—
8	Number of multiple-choice items	4	1	7	5	6	3	4	4	3	4	—
	Number of short constructed-response items	5	1	3	5	6	1	1*	7*	8	0	—
	Number of regular constructed-response items	0	5	0	0	0	3	4	0	0	7	—
	Number of extended constructed-response items	1	1	1	2	1	2	0	1	1	2	—
	Total	10	8	11	12	13	9	9	12	12	13	—
12	Number of multiple-choice items	4	3	1	5	5	1	4	4	7	10	0
	Number of short constructed-response items	5	5	0	5	7*	0	1*	7*	7	4	0
	Number of regular constructed-response items	0	0	6	0	0	6	4	0	0	0	6
	Number of extended constructed-response items	1	1	1	2	0	1	0	1	0	2	1
	Total	10	9	8	12	12	8	9	12	14	16	7

**__Note__: This block contained one additional extended constructed-response item. The categories of that item were combined such that the version used in scaling had only two categories (see Section 12.4.3.1.2). Therefore, the item is counted as "short constructed-response," reflecting the amount of information it contributed to the final scale.*

BEST COPY AVAILABLE

Table 12-4

Numbers of Scaled Reading Items Common Across Grade Levels

Grade	Reading for Literary Experience	Reading to Gain Information	Reading to Perform a Task	Total
4 only	32	31	0	63
8 only	8	25	12	45
12 only	17	43	14	74
4 and 8 only	11	10	0	21
8 and 12 only	10	12	21	43
4 and 12 only	0	0	0	0
4, 8, and 12	0	0	0	0
Total	78	121	47	246

BEST COPY AVAILABLE

Table 12-5

Descriptive Statistics for Item Blocks by Position Within Booklet and Over All Occurrences for the Reading Main Sample, Grade 4

		Block							
		Reading for Literary Experience				Reading to Gain Information			
Statistic		**R3**	**R4**	**R5**	**R9**	**R6**	**R7**	**R8**	**R10**
Number of scaled items	–	11	12	11	9	10	10	9	12
Unweighted sample size	1	924	913	898	894	925	911	940	928
	2	936	899	899	938	898	910	922	922
	ALL	1,860	1,812	1,797	1,832	1,823	1,821	1,862	1,850
Average weighted item score	1	.61	.64	.45	.62	.58	.43	.53	.66
	2	.58	.63	.43	.58	.54	.42	.50	.61
	ALL	.60	.63	.44	.60	.56	.42	.51	.64
Average weighted r-polyserial	1	.71	.72	.63	.57	.61	.68	.63	.66
	2	.78	.72	.64	.69	.65	.71	.65	.68
	ALL	.75	.72	.63	.64	.63	.69	.64	.67
Weighted alpha reliability	1	.80	.82	.74	.73	.70	.74	.73	.79
	2	.83	.83	.74	.79	.72	.75	.76	.78
	ALL	.82	.82	.74	.76	.71	.74	.74	.78
Weighted proportion of students attempting last item	1	.76	.61	.73	.63	.70	.61	.69	.81
	2	.85	.74	.83	.78	.82	.70	.79	.87
	ALL	.80	.67	.78	.70	.76	.65	.74	.84

Table 12-6

Descriptive Statistics for Item Blocks by Position Within Booklet and Over All Occurrences for the Reading Main Sample, Grade 8

Statistic	Block Position	Block									
		Reading for Literary Experience			Reading to Gain Information				Reading to Perform a Task		
		R3	R4	R5	R6	R7	R8	R13*	R9	R10	R11
Number of scaled items	—	10	8	11	12	13	9	13	9	12	12
Unweighted sample size	1	929	931	922	899	923	934	—	909	899	918
	2	920	911	914	906	923	912	—	883	921	896
	ALL	1,849	1,842	1,836	1,805	1,846	1,846	1,820	1,792	1,820	1,814
Average weighted item score	1	.46	.42	.65	.57	.67	.56	—	.60	.61	.67
	2	.44	.38	.66	.51	.65	.52	—	.59	.57	.66
	ALL	.45	.40	.66	.54	.66	.55	.65	.60	.58	.67
Average weighted r-polyserial	1	.67	.64	.74	.65	.69	.60	—	.68	.62	.66
	2	.68	.68	.72	.66	.72	.64	—	.69	.64	.72
	ALL	.68	.66	.73	.66	.71	.62	.62	.69	.63	.69
Weighted alpha reliability	1	.74	.70	.78	.73	.78	.71	—	.71	.75	.77
	2	.75	.76	.77	.75	.82	.73	—	.72	.76	.80
	ALL	.75	.73	.78	.74	.80	.72	.76	.71	.76	.78
Weighted proportion of students attempting last item	1	.73	.66	.94	.83	.81	.89	—	.94	.78	.80
	2	.81	.61	.95	.85	.84	.93	—	.94	.81	.90
	ALL	.77	.64	.94	.84	.82	.91	.93	.94	.80	.85

****Note:*** *A 50-minute block that comprised an entire booklet.*

Table 12-7

Descriptive Statistics for Item Blocks by Position Within Booklet and Over All Occurrences for the Reading Main Sample, Grade 12

Statistic	Block Position	Block										
		Reading for Literary Experience			Reading to Gain Information					Reading to Perform a Task		
		R3	R4	R5	R6	R7	R8	R13*	R14*	R9	R10	R11
Number of scaled items	—	10	9	8	12	12	8	16	7	9	12	14
Unweighted sample size	1	742	762	760	793	749	750	—	—	766	742	781
	2	746	713	743	750	766	713	—	—	788	794	761
	ALL	1,488	1,475	1,503	1,543	1,515	1,463	1,514	1,499	1,554	1,536	1,542
Average weighted item score	1	.60	.54	.45	.67	.52	.59	—	—	.74	.71	.52
	2	.59	.48	.43	.65	.51	.56	—	—	.74	.68	.54
	ALL	.60	.51	.44	.66	.52	.57	.61	.42	.74	.70	.53
Average weighted r-polyserial	1	.73	.64	.64	.65	.57	.62	—	—	.72	.63	.58
	2	.72	.66	.65	.71	.58	.65	—	—	.75	.66	.59
	ALL	.72	.65	.65	.68	.57	.64	.62	.68	.74	.65	.58
Weighted alpha reliability	1	.78	.63	.72	.66	.59	.69	—	—	.67	.72	.70
	2	.78	.65	.72	.72	.58	.72	—	—	.73	.74	.70
	ALL	.78	.64	.72	.69	.59	.70	.79	.70	.70	.73	.70
Weighted proportion of students attempting last item	1	.83	.60	.77	.88	.79	.86	—	—	.96	.85	.99
	2	.82	.72	.79	.89	.82	.90	—	—	.94	.87	1.00
	ALL	.83	.66	.78	.88	.80	.88	.94	.95	.95	.86	1.00

**Note*: *This was a 50-minute block that comprised an entire booklet.*

Small but consistent differences were noted based upon whether a block appeared first or second within a booklet. When the block appeared first in the booklet, the average item score tended to be higher and the average polyserial correlation tended to be lower. The largest differences were noted in the proportion of students not attempting the last item in the block; more students attempted the last item when the block appeared in the second position. It appears that students learned to pace themselves through the second block, based on their experience with the first block. Recall that the design of the reading assessment is not completely balanced. Thus, when these serial position effects were first noticed, it was feared that they might adversely affect the results of the IRT scaling. As part of the analysis of the 1992 reading assessment, a special study was completed to examine the effects of the serial position differences. The serial position effects were found to have minimal results on the scaling, most likely due to the balance of the partial BIB design of the booklets. The effects portrayed in Tables 12-5 through 12-7 are similar in size to the effects observed in the 1992 reading assessment, and were therefore, unlikely to produce adverse effects on the final IRT scaling.

12.3.2 Scoring the Constructed-Response Items

As indicated earlier, the reading assessment included constructed-response items. Responses to these items were included in the scaling process. In addition, detailed analyses of the ordinal responses to the constructed-response items were also conducted, and are summarized in the *NAEP 1994 Reading Report Card for the Nation and the States: Findings from the National Assessment of Educational Progress and Trial State Assessments*. Chapter 7 provides the ranges for percent agreement between raters for the items as they were originally scored. The percent agreement for the raters and Cohen's (1968) Kappa are given in Appendix H.

12.3.3 Differential Item Functioning

Prior to scaling, differential item functioning (DIF) analyses were conducted on the reading data for grades 4, 8, and 12, and for the Trial State Assessment sample at grade 4. The purpose of these analyses was to identify items that were differentially difficult for various subgroups and to reexamine such items to determine their fairness and the appropriateness of including them in the scaling. The information in this section focuses mainly on the analyses conducted on the national data; results for the analyses of data from the Trial State Assessment are described in the technical report for that assessment.

DIF analyses were based upon the modification of the Mantel-Haenszel procedure as adapted by Holland and Thayer (1988) and its extension for polytomous items described in Zwick, Donoghue, and Grima (1993). DIF procedures are described in Chapter 9. For multiple-choice and short constructed-response items (i.e., dichotomous items), DIF decisions were based upon the Mantel-Haenszel procedure. For regular and extended constructed-response items, the extended procedure (Mantel, 1963) was used.

The "grade-only" portion of the main focused-BIB was used for DIF analyses. Sample sizes were large enough to compare male and female students, White and Black students, and White and Hispanic students. Weights were rescaled separately for each comparison, as described in Chapter 9. DIF analyses were conducted separately by grade. A given item was subjected to at least three, and as many as nine, separate DIF analyses.

For dichotomous items, the DIF index generated by the Mantel-Haenszel procedure is commonly used by ETS to place items into one of three categories: "A," "B," or "C." "A" items exhibit little or no DIF, while "C" items exhibit a strong indication of DIF and should be examined more closely. Positive values of the index indicate items that are differentially easier for the "focal" group (female, Black, or Hispanic students) than for the "reference" groups (male or White students). Similarly, negative values indicate items that are differentially harder for the focal group than the reference group. An item that was classified as a "C" item in *any* analysis was considered to be a "C" item. Sixteen "C" items were identified in the main reading assessment. Table 12-8 summarizes the results of DIF analyses for dichotomously scored items.

For polytomous items (regular constructed-response and extended constructed-response), a similar scheme was developed to classify items (this is discussed in detail in Donoghue, 1995). Polytomous items were placed into two categories, "AA" or "CC." "AA" items exhibit no DIF, while "CC" items exhibit a strong indication of DIF and should be examined more closely. The classification criterion for polytomous items is presented in Donoghue (1995). As with dichotomous items, positive values of the index indicate items that are differentially easier for the "focal" group (female, Black, or Hispanic students) than for the reference groups (male or White students). Similarly, negative values indicate items that are differentially harder for the focal group than the reference group. An item that was classified as a "CC" item in *any* analysis was considered to be a "CC" item. Sixteen "CC" items were identified in the main reading assessment. Table 12-9 summarizes the results of DIF analyses for polytomously scored items.

Following standard practice at ETS for DIF analyses conducted on final forms, all "C" and "CC" items were reviewed by a committee of trained test developers and subject-matter specialists. As described in Chapter 9, such committees are charged with making judgments about whether or not the differential difficulty of an item is *unfairly* related to group membership. The committee assembled to review NAEP items included ETS staff and outside members with expertise in the field. The committee carefully examined each "C" and "CC" item to determine if either the language or contents would tend to make the item more difficult for an identified group of examinees. It was the committee's judgment that none of the "C" or "CC" items for the national or the Trial State Assessment data were functioning differentially due to factors irrelevant to test objectives. Hence, none of the items were removed from scaling due to DIF.

Table 12-8

DIF Category by Grade for Dichotomous Items

Grade	DIF Category*	Analysis		
		Male/Female	**White/Black**	**White/Hispanic**
4	C-	1	0	0
	B-	1	7	5
	A-	32	36	32
	A+	33	21	29
	B+	0	3	1
	C+	0	0	0
8	C-	1	0	0
	B-	5	2	6
	A-	35	30	27
	A+	33	36	42
	B+	2	9	2
	C+	1	0	0
12	C-	2	1	0
	B-	5	5	4
	A-	42	28	34
	A+	32	43	45
	B+	3	8	2
	C+	1	0	0

***Note:** Positive values of the index indicate items that are differentially easier for the "focal" group (female, Black, or Hispanic students) than for the "reference" groups (male or White students). "A" means no indication of DIF. "B" means a weak indication of DIF, and "C" means a strong indication of DIF.*

Table 12-9

DIF Category by Grade for Polytomous Items

Grade	DIF Category	Analysis		
		Male/Female	White/Black	White/Hispanic
4	CC-	0	0	0
	AA-	4	10	11
	AA+	13	7	6
	CC+	0	0	0
8	CC-	0	0	0
	AA-	9	17	13
	AA+	23	15	19
	CC+	0	0	0
12	CC-	0	0	0
	AA-	13	18	16
	AA+	22	17	19
	CC+	0	0	0

**Note: Positive values of the index indicate items that are differentially easier for the "focal" group (female, Black, or Hispanic students) than for the "reference" groups (male or White students). "AA" means little or no indication of DIF and "CC" means a strong indication of DIF.*

12.4 IRT SCALING

12.4.1 Overview of Item Parameter Estimation

In 1992, separate IRT-based scales were developed for each of the purposes for reading identified in the reading framework. As described in Chapter 11, multiple-choice items were fit using a 3PL model. Short constructed-response items were fit using a 2PL model. Regular and extended constructed-response items were fit using the generalized partial credit model.

For calibration, all items that were not reached were treated as if they had not been presented to the examinees[3]. Recall that responses to regular and extended constructed-response items that were off-task were also treated as if they had not been presented. The treatment of omitted responses differed according to the item type. Omitted responses to multiple-choice items were treated as fractionally correct (see Section 11.3.1 and Mislevy & Wu, 1988, for a discussion of these conversions). Omitted responses to short constructed-response items were treated as incorrect, and omitted responses to regular and extended constructed-response items were assigned to the lowest category.

---

[3] An exception to this rule was the treatment of extended constructed-response items at the end of the block. See Section 12.3.1 for a discussion.

For each purpose of reading, three separate scalings, one for each age/grade sample, were conducted. The analyses were conducted on the following samples:

- the 1992 age 9/grade 4 sample with the 1994 age 9/grade 4 national sample;
- the 1992 age 13/grade 8 sample with the 1994 age 13/grade 8 national sample; and
- the 1992 age 17/grade 12 sample with the 1994 age 17/grade 12 national sample.

Item parameters were estimated using combined data from both assessment years. Items that were administered for more than one assessment (trend items) were constrained to have equal item response functions across assessment years. However, some items exhibited clear evidence of functioning differently across assessments (see discussion in Section 12.4.3.2). These items were treated as separate items for each assessment year.

The calibration was performed using all the available examinees. Student sampling weights were used for the analysis. For scaling, sampling weights were restandardized to ensure that each assessment year had a similar sum of weights, and so had approximately equal influence in the calibration. Each assessment year's data were treated as a sample from a separate subpopulation. Thus, separate proficiency distributions were estimated for each assessment year.

Item responses were calibrated using the BILOG/PARSCALE program. Starting values were computed from item statistics based on the entire data set. BILOG/PARSCALE calibrations were done in two stages. At stage one, the proficiency distribution of each assessment year was constrained to be normally distributed, although the means and variances differed across assessments. The values of the item parameters from this normal solution were then used as starting values for a second-stage estimation run in which the proficiency distribution (modeled as a separate multinomial distribution for each assessment) was estimated concurrently with item parameters. Calibration was concluded when changes in item parameters became negligibly small.

A complexity introduced by the 50-minute blocks in reading is that those blocks of items must be linked in some way to the shorter blocks. This is complicated by the fact that no students received the shorter blocks in addition to the 50-minute blocks. Because the samples of students receiving each booklet are representative of the population as a whole, it was assumed that the distribution of student proficiency was the same for the students receiving the 50-minute blocks as for the students receiving the booklets containing the shorter blocks.

Two factors complicated the scaling process for the short-term trend from 1992 to 1994. The first factor involved the treatment of omitted responses to polytomously scored (regular constructed-response and extended constructed-response) items. This necessitated the reanalysis of the 1992 reading assessment, which is discussed in Section 12.4.2. The second complication arose from modifications to process of rating polytomously scored items, specifically in the training of raters and the introduction of the image-based scoring. These changes led to adjustments in how the scoring rubrics were applied to some short constructed-response items. This is discussed in Section 12.4.3.1.

12.4.2 Reanalysis of the 1992 Reading Assessment

As noted above (and described in Chapter 9), a distinction is made between missing responses at the end of each block (not reached) and missing responses prior to the last observed response (omitted). Items that were not reached were treated as if they had not been presented to the examinee. In IRT scaling, omitted responses to multiple-choice items are treated as "fractionally correct" (see Chapter 11). Omitted responses to short constructed-response items (dichotomously scored) were regarded as incorrect. For polytomously scored items, omitted responses were combined with incorrect (lowest category) responses.

During the analysis of the 1994 reading assessment, it was discovered that, due to a computer documentation error, BILOG/PARSCALE did not treat omitted responses to polytomous items as intended. Rather than being combined with the lowest valid response category, omitted responses were actually treated as if the item had not been presented.

The error was corrected and the data from the 1992 assessment were reanalyzed using the procedures described in Donoghue, Isham, Bowker, and Freund (1994). All decisions regarding treatment of items in scaling were reexamined and accepted. The conditioning model was reestimated using the CGROUP program, and new plausible values were generated. As in the original analysis, for each purpose for reading, scale means and standard deviations were set to 250.0 and 50.0 using the transformation:

$$\theta_{proficiency} = A \bullet \theta_{calibrated} + B$$

where $\theta_{proficiency}$ denotes values on the final transformed scale and $\theta_{calibrated}$ denotes values on the original calibration scale from BILOG/PARSCALE.

For the Reading for Literary Experience and Reading to Gain Information scales, the linear indeterminacy was resolved by transforming the mean and standard deviation of the three age/grade samples combined together to the 250.0, 50.0 metric. The third scale, Reading to Accomplish a Task, was only presented at age 13/grade 8 and age 17/grade 12. For this scale, the means and standard deviations of each of these two age/grade samples were matched to the average of the other two scales to determine the appropriate transformation. The constants for the linear transformation for each scale are given in Table 12-10.

Table 12-10

Reanalysis of 1992 Reading Assessment:
Coefficients of Linear Transformations of the Scales
from the Calibrating Scale Units to the Units of Reporting Proficiency

Scale	A	B
Reading for Literary Experience	50.32	247.73
Reading to Gain Information	50.26	256.25
Reading to Perform a Task	44.49	270.70

Extensive comparisons between the original results and the results of the reanalysis revealed only minor discrepancies of little substantive importance; the two sets of results were highly linearly related ($r > .99$ for each age/grade sample). Revised results for the 1992 assessment are presented in *1994 NAEP Reading: A First Look—Findings from the National Assessment of Educational Progress* (Williams, Reese, Campbell, Mazzeo, & Phillips, 1995) and in the *NAEP 1994 Reading Report Card for the Nation and the States: Findings from the National Assessment of Educational Progress and Trial State Assessments* (Campbell, Donahue, Reese, & Phillips, 1995).

12.4.3 Evaluation of Model Fit

During and subsequent to item parameter estimation, evaluations of the fit of the IRT models were carried out for each of the items. These evaluations were based primarily on graphical analysis. First, model fit was evaluated by examining plots of nonmodel-based estimates of the expected proportion correct (conditional on proficiency) versus the proportion correct predicted by the estimated item response function (see Chapter 9 and Mislevy & Sheehan, 1987, p. 302). Figure 12-1 gives an example plot of a multiple-choice item that demonstrates good model fit, R016102, from the Reading to Accomplish a Task scale at age 13/grade 8. For regular and extended constructed-response items, similar plots were produced for each item category response function (see Chapter 9). Figure 12-2 gives an example plot of an extended constructed-response item that demonstrates good model fit, R015804, from the Reading for Literary Experience scale at age 9/grade 4. Note that the remaining item plots in this section (Figures 12-3 through 12-8) were obtained from preliminary item parameter calibrations. They are presented to reflect the information used to make the decisions discussed in the text. Plots produced from the final item parameters (listed in Appendix D) were very similar to those presented and supported the decisions made.

12.4.3.1 Items Deleted from the Final Scale

In making decisions about excluding items from the final scales, a balance was sought between being too stringent, hence, deleting too many items and possibly damaging the content representativeness of the pool of scaled items, and being too lenient, hence including items with model fit poor enough to endanger the types of model-based inferences made from NAEP results. For the majority of the items, the model fit was extremely good. Items that clearly did not fit the model were not included in the final scales; however, a certain degree of misfit was tolerated for a number of items included in the final scales.

At age 17/grade 12, one of the new (unique to 1994) items from the Reading to Gain Information scale, R016603, was dropped from the final scales due to poor fit to the IRT model. Figure 12-3 gives an IRT plot of this item. For proficiency values above -1.0, categories 0 and 1 provide virtually no discrimination; the empirical item category response functions are essentially flat. Nor did collapsing categories improve the measurement properties of the item. Thus, the item was deleted.

In addition, two short-term trend items were deleted for 1994 only (i.e., these items were treated as if they were unique to the 1992 assessment). One short constructed-response item (R013108, administered at both age 13/grade 8 and age 17/grade12) showed evidence of item parameter drift. R013108 was an extremely difficult item in 1992, and as scored in 1994, the probability of a correct response was too small to allow stable estimates of item parameters for 1994 (P+ = .001 at age 13/grade 8, P+ = .006 at age 17/grade 12).

The second deleted short-term trend item, R013915, was an extended constructed-response item. It was presented only at age 17/grade 12. In 1994, a very large percentage of the student responses were missing (63%). This was a large increase from 25 percent in 1992. The reason for the increase across assessments is not fully understood at present. Given the large percentage of students who were missing data for this item in 1994, it could not be reliably included in the scaling. Thus, the item was dropped. Table 12-11 gives the items that were deleted from the 1994 scaling.

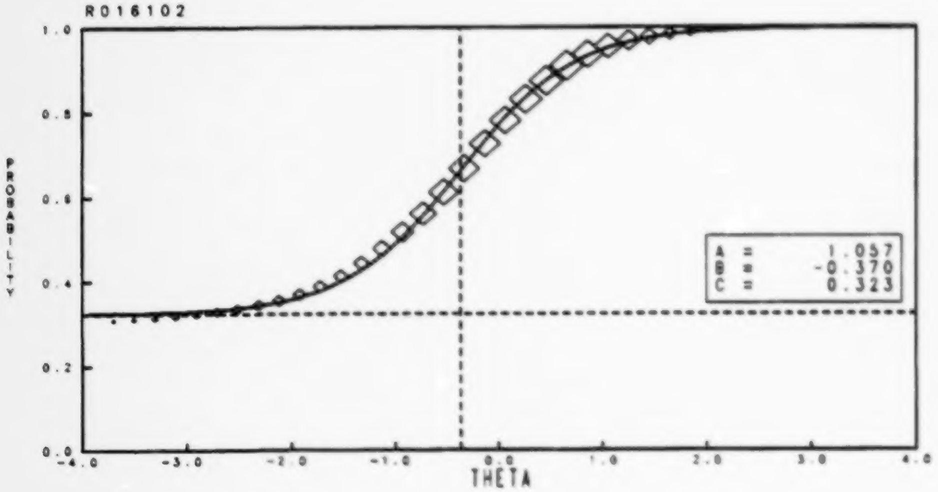
12.4.3.1.1 Recoded Polytomous Items

Polytomous items received special treatment (i.e., recoding) for one of two reasons. First, some of the short-term trend items were recoded in the original 1992 scaling. These items were recoded again for the 1994 assessment. Second, two of the new (unique to 1994) polytomous items received special treatment in the scaling. Figure 12-4 shows one such item, R015707 from the Reading for Literary Experience scale at age 9/grade 4. There is a lack of fit for both the unsatisfactory and partial categories for low proficiency ($\theta < -2.0$) values. There is also a marked misfit for categories 2 and 3 in the range ($1.0 < \theta < 2.0$). Categories 2 and 3 of this item were collapsed:

0 = Unsatisfactory
1 = Partial
2 = Essential
3 = Extensive

Figure 12-5 shows the recoded version of R015707 from the final scaling. The fit is substantially improved for the low ability examinees. However, there is still some misfit for $\theta > 1.0$. Table 12-12 lists polytomous items that were recoded for scaling in 1994.

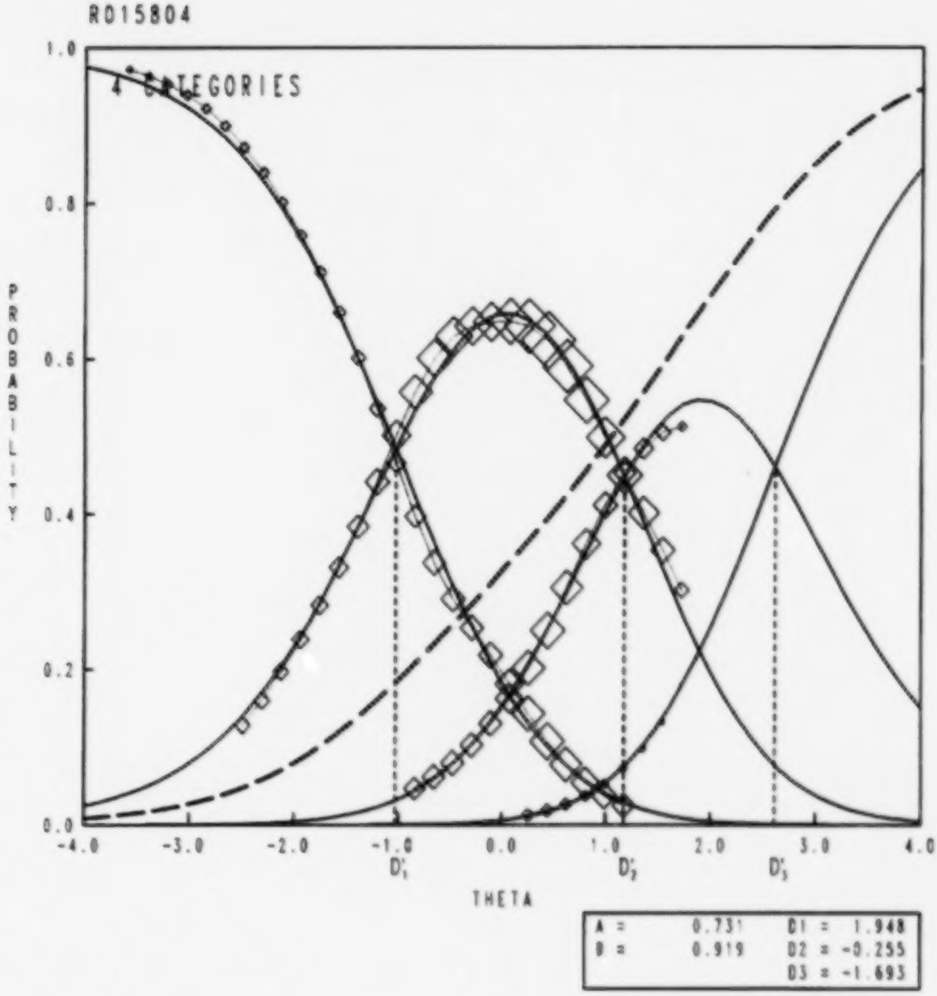
Figure 12-1
Example Cross-Sectional Dichotomous Item (R016102, Age 13/Grade 8)
*Exhibiting Good Model Fit**



***Note**: The plot compares empirical and model-based estimates of the item response function (IRF). The smooth curve represents the model-based estimate at each provisional proficiency level. The diamonds represent 1994 age 13/grade 8 data.*

Figure 12-2

*Example Cross-Sectional Polytomous Item (R015804, Age 9/Grade 4) Exhibiting Good Model Fit**

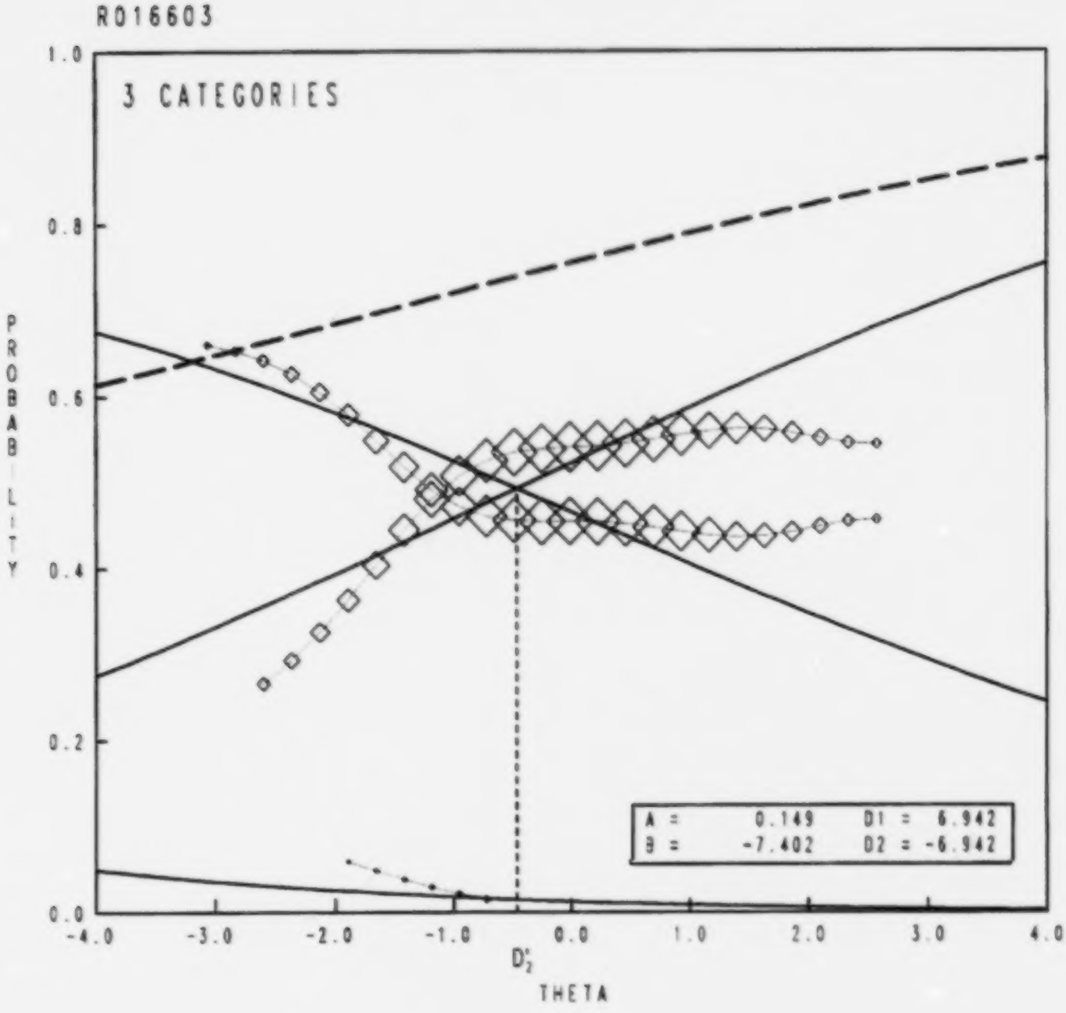


****Note:*** *The plot compares empirical and model-based estimates of the item category response functions (ICRFs). The smooth curve represents the model-based estimate at each provisional proficiency level. The diamonds represent 1994 age 9/grade 4 data.*

BEST COPY AVAILABLE

Figure 12-3

Example Cross-Sectional Polytomous Item (R016603, Age 17/Grade 12)
*Exhibiting Unacceptably Poor Model Fit**



****Note:*** *The plot compares empirical and model-based estimates of the item category response functions (ICRFs). The smooth curve represents the model-based estimate at each provisional proficiency level. The diamonds represent 1994 age 17/grade 12 data.*

Table 12-11

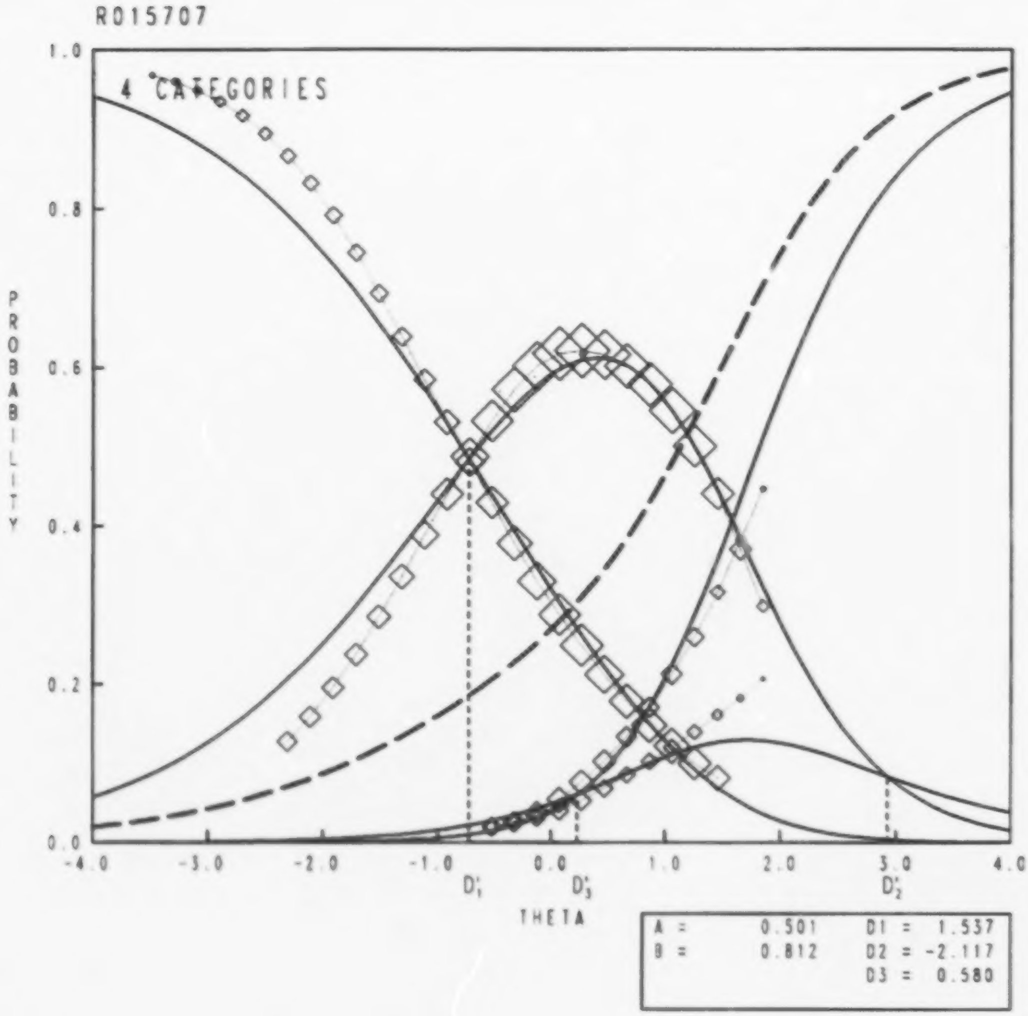
Items Deleted from the Final Scaling

Scale	NAEP ID	Block	Grade(s) Affected	Reason for Decision
Reading for Literary Experience	R013108*	R3	8, 12	Extremely low P+
Reading to Gain Information	R016603	R14	8, 12	Poor fit in 1994
Reading to Perform a Task	R013915*	R11	12	Large number of omit

*__*Note*__*: Trend items used in 1992 but deleted in 1994.*

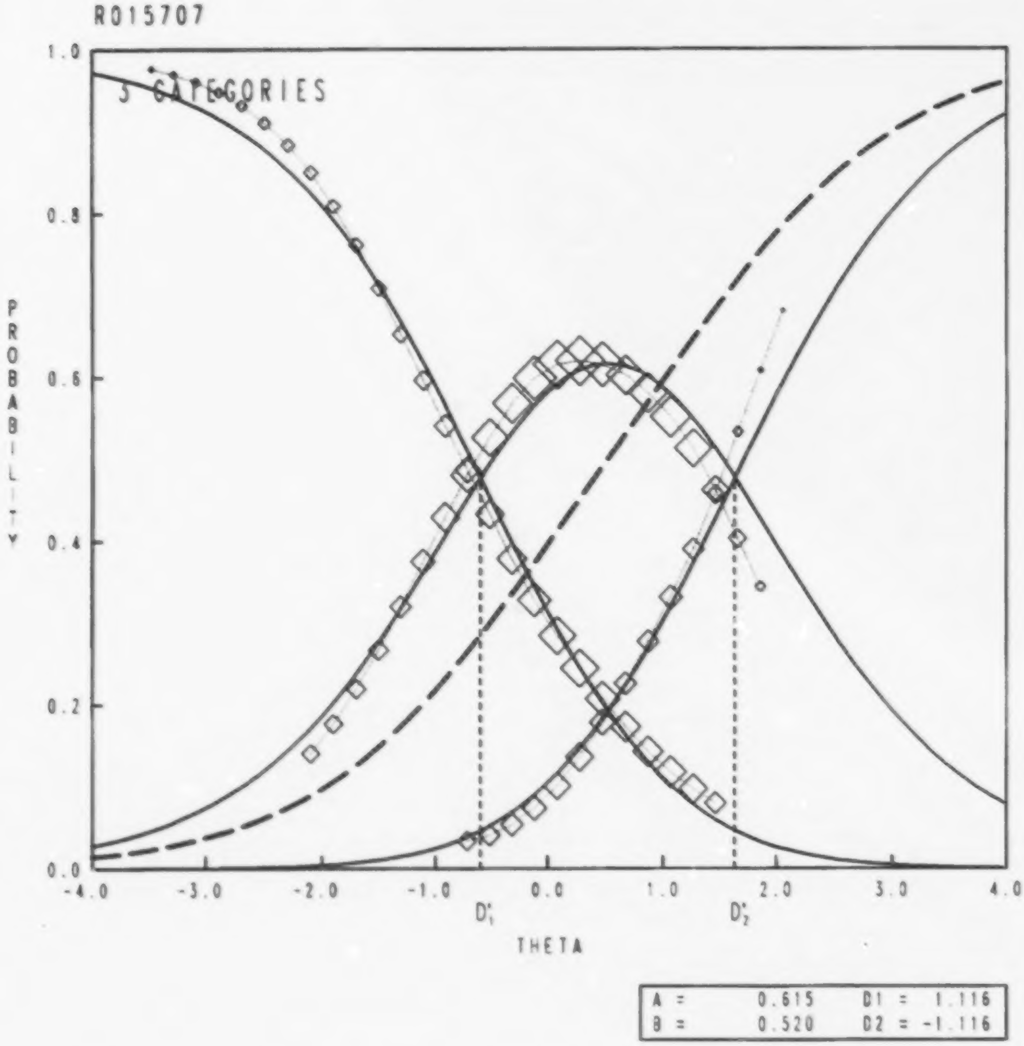
Figure 12-4

Example Cross-Sectional Polytomous Item (R015707, Age 9/Grade 4)
*Exhibiting Poor Model Fit**



****Note:*** *The plot compares empirical and model-based estimates of the item category response functions (ICRFs). The smooth curve represents the model-based estimate at each provisional proficiency level. The diamonds represent 1994 age 9/grade 4 data.*

Figure 12-5
Example Cross-Sectional Polytomous Item (R015707, Age 9/Grade 4)
*After Collapsing Categories 2 and 3**



****Note:*** *The plot compares empirical and model-based estimates of the item category response function (ICRF). The smooth curve represents the model-based estimate at each provisional proficiency level. The diamonds represent 1994 age 9/grade 4 data.*

BEST COPY AVAILABLE

Table 12-12

Recoding of Polytomous Items for Scaling

Scale	NAEP ID	Block	Grade(s) Affected	Reason for Decision	Disposition
Reading for Literary Experience	R012111	R4	4	Recoded in 1992	Combine categories 0 + 1
	R013506	R4	12	Recoded in 1992	Combine categories 0 + 1
	R013610	R5	12	Recoded in 1992	Combine categories 0 + 1, 2 + 3 (dichotomize)
Reading to Gain Information	R015707	R8	4	Poor fit in 1994	Combine categories 2 + 3
	R013706	R7	12	Recoded in 1992	Combine categories 0 + 1, 2 + 3 (dichotomize)
	R013805	R8	12	Recoded in 1992	Combine categories 0 + 1
Reading to Perform a Task	R013004	R11	8	Recoded in 1992	Combine categories 0 + 1
	R013404	R10	8, 12	Recoded in 1992	Combine categories 0 + 1
	R013406	R10	8, 12	Recoded in 1992	Combine categories 0 + 1, 2 + 3 (dichotomize)
	R016104	R9	8, 12	Poor fit in 1994	Combine categories 1 + 2 (dichotomize)

12.4.3.2 Fit Across of a Common Item Category Response Function Across Assessment Years

The adequacy of the assumption of a common item response function across assessment years was also evaluated. For dichotomous items, this was evaluated by comparing the nonmodel-based expected proportions for each assessment year to the single, model-based item response function fit by BILOG/PARSCALE. For polytomously scored items, similar plots were produced for each item category response function (ICRF, see Chapter 9). Plots showing each assessment year's data separately and the common item category response function were then examined. Items that showed clear evidence of functioning differently across assessments were treated as separate items for each assessment year. As was the case with deleting items, in making decisions about scaling items separately by assessment year, a balance was sought between being too stringent, hence, splitting too many items and possibly damaging the common item link between the assessment years, and being too lenient, hence, including items with model fit poor enough to endanger the model-based trend inferences.

At each grade, several items were calibrated separately for each assessment year for one of two reasons. First, examination of residual plots identified a few items as functioning differently across assessments. Second, as noted above, changes in scoring procedures caused a number of short constructed-response items (dichotomously scored) to exhibit item parameter drift. These items were calibrated separately across assessments due to rater drift across assessment years.

12.4.3.2.1 Short-Term Trend Scorer/Item Parameter Drift

For the 1994 assessment, several changes were made to the scoring procedures of the short-term trend (1992) items. Scorer training materials were improved to remove potential ambiguities in the scoring rubrics of some items. The scoring process was also modified. In 1992, scoring used intact examinee booklets; scorers would score all short constructed-response items in a given booklet before proceeding to the next booklet. Because booklets were not scored in any systematic order, this procedure effectively required scorers to retain the rubrics to all of the short constructed-response items in the assessment. These procedures applied only to the scoring of the short constructed-response items; in 1992 the extended constructed-response items were scored one item at a time to ensure adequate interrater reliability.

In 1994, NAEP introduced an image-based scoring system (see Chapter 6). Each response to a constructed-response item was electronically scanned and presented to raters as a video image for scoring. The image-based system allowed raters to focus on a single item at a time. Raters received training on one item, then scored large numbers of student responses to that item before proceeding to the next item. There is evidence from the 1993 field test (Johnson & Freund, 1994) that the image-based system *per se* had little effect on the scoring. This finding was confirmed by an additional rescore study performed within the 1994 reading assessment. However, the reduction in the cognitive demand placed upon the raters may well have altered how the scoring rubrics were applied.

The improvements in training materials and movement to image-based scoring improved the rigor with which the scoring rubrics were applied, and hence the overall quality of the rating of student responses. Examination of ratings assigned to individual responses confirmed the general superiority of the 1994 procedures. However, a clear danger with *any* change in scoring procedures is that it may result in item parameter drift; that is, the ratings from the two sets of procedures may differ in their relationship to the underlying trait. In fact, early item analyses revealed that several short constructed-response items showed implausibly large changes (when compared to other item types) in the percentage of examinees who got the item correct.

Identifying item parameter drift can be difficult; at this time there are no simple statistics with known distributions to use for a statistical test of item parameter drift. It is important to separate differences in proficiencies of the two assessments from the functioning of the item. Simply comparing the probability of a correct response (p-values) is not a satisfactory procedure for assessing drift. The proportion correct on an item is a valid measure of drift only if the examinees in the two assessments have identical proficiency distributions. When the assessments differ in proficiency, comparison of p-values confounds drift with impact. To accurately assess drift, it is necessary to condition on some measure of proficiency. However, the estimate of proficiency can itself be affected by the presence of item parameter drift. It is important that the items used in estimating the proficiency distributions of the two assessments be, as nearly as possible, free of item parameter drift. The process of identifying and removing drift items is often termed "purification." (Lord, 1980)

Because of the complexity in identifying item parameter drift, a two-stage procedure was used to identify short constructed-response items that exhibited item parameter drift. First, based on rescore data, a number of short constructed-response items were tentatively identified as exhibiting drift. These items were temporarily excluded from calibration, and an initial estimate of the ability distribution for each assessment was obtained. Next, IRT diagnostic item plots were examined, and items exhibiting item parameter drift were identified. Items that demonstrated item parameter drift were split (treated as two separate items, one administered only in the 1992 assessment and one administered only in 1994). Sections 12.4.3.2.1.1 and 12.4.3.2.2 describe the identification of drift items in more detail.

12.4.3.2.1.1 Preliminary Screening Based on Rescore Data

For each short-term trend constructed-response item, a sample of approximately 500-600 of the 1992 responses was rescored in 1994 using the new procedures. All items showed an acceptably high level of exact agreement. However, several items showed a clear trend in the disagreements. For example, consider the rescore data for a particular item. For the papers that were rescored, two sets of ratings are available—one set from the 1992 raters and another set from the 1994 raters. Of those papers for which 1992 and 1994 raters disagreed, many responses might have been deemed acceptable by the 1992 raters but not acceptable by the 1994 raters. On the other hand, relatively few responses were rated acceptable in 1994 but not in 1992. Overall, more responses were deemed acceptable in 1992, and the net effect of such a pattern would be that the item is more difficult as it was scored in 1994.

To detect such drifts in difficulty, a paired t-test was computed for the rescored sample on each short constructed-response item. The two ratings (original 1992 and rescored in 1994) served as the dependent variable. No design effect was used in this test. Any item for which the t-test was significant ($p \leq .05$) was preliminarily identified as exhibiting item parameter drift.

This initial screening was intentionally liberal in identifying potential drift items. No modification of the test was made to account for clustering in the data, or was any adjustment made to account for multiple hypothesis testing. The point of the initial screening was to purify the test of possible drift items prior to IRT calibration and final determination of which items exhibited item parameter drift (i.e., Lord, 1980). Final determination of whether or not an item drifted was based on examination of IRT-based model fit plots, as discussed in the next section.

12.4.3.2.1.2 Identification of Drift Items Based on IRT Model-Fit Plots

An initial IRT calibration was conducted with every potential drift item (identified using the rescore data) excluded from the calibration. All other short-term trend items in the assessment were assumed to function identically across the two assessment years. BILOG/PARSCALE estimated a separate ability distribution for each assessment year. Each assessment year's distribution was assumed to be normally distributed, but the mean and standard deviation was allowed to vary across assessment years. The overall, total population distribution was constrained to have a mean of zero and unit variance.

The prior distribution was then fixed for each year. This estimate was purified of the most severe drift items, and reflected any shifts in reading proficiency across the assessments. Using this fixed prior, a common (identical across assessment years) item category response function was fit for each potential drift item. Model fit of dichotomous items was evaluated by examining plots of nonmodel-based estimates of the expected proportion correct (conditional on proficiency) versus the proportion correct predicted by the estimated item response function. Items that showed evidence of drift (i.e., clear separation between the two years' data) were then split.

The identification of drift items was an iterative process. At each stage, the worst items (i.e., those showing the most evidence of drift) were split; a separate item category response function was fit to each assessment year's data. Item parameters and ability distributions were reestimated. Finally, a new set of item plots was generated. This three-step procedure was repeated until no new items were identified.

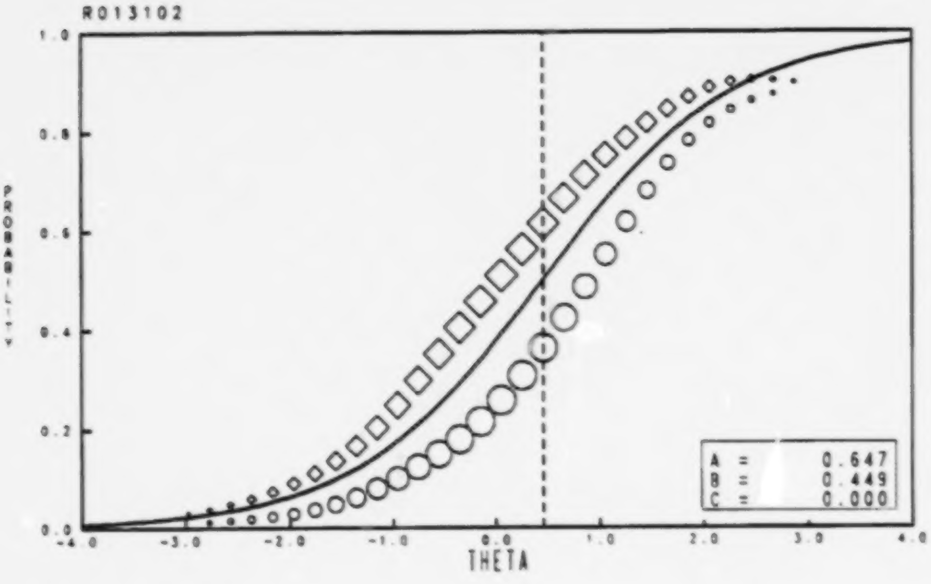
Figure 12-6 gives an example item plot for an item that was split early in the process, R013102 at age 13/grade 8. The ovals represent data from the 1992 assessment, and the diamonds represent the data from the 1994 assessment. There is a marked separation between the two sets of symbols that indicate that the item was substantially more difficult as scored in 1994. Figure 12-7 shows the result of splitting this item. Figure 12.7a gives the IRF fit using only the 1992 data, and Figure 12-7b gives the IRF fit to the 1994 data. Within each assessment, there is good agreement between the curve and the plotted points.

Figure 12-8 shows another item that yielded a significant *t*-test – R013405 which was administered at age 17/grade 12. Here the symbols representing the 1992 and 1994 data are not well separated. Although there are some discrepancies between the points for the two assessment years, the differences from the common IRF are generally small (< 0.1). This item was *not* split between assessments.

Tables 12-13 through 12-15 list the short-term trend items that were calibrated separately across assessment years. The large number of items treated separately at age 17/grade 12 raised questions about the validity/stability of the remaining items as a measure of trend. To investigate this issue, a study was conducted using the 1992 age 17/grade 12 reading data. The purpose of the study was to approximate the effect of applying the more appropriate 1994 scoring rubrics to the 1992 data. The study used the items identified as questionable by the analysis of the rescore data. For each item, a 2 x 2 table was computed based upon the rescore. Let X_j represent the score an item received as scored in 1992. Similarly, let Y_j represent the score a response to item *j* received in the 1994 scoring. For each item, the conditional probabilities of the 1994 score, given the 1992 score, was computed: that is, $P(Y_j=y|X_j=x)$. Next, the response to each 1992 item was examined. If the response was correct, it was allowed to remain with probability $P(Y_j=1|X_j=0)$ and changed to incorrect with probability $P(Y_j=0|X_j=1)$. Incorrect responses were modified similarly. The resulting data approximated the effect of applying the more appropriate 1994 scoring rubrics to the 1992 data. The 1992 data were then rescaled and reconditioned, and the results compared to those of the original 1992 analysis. The comparison indicated that the original and modified data yielded results that were essentially a linear transformation of one another; all remaining differences were less than two standard errors from the original results. Thus, it was concluded that the remaining (nondrift) items measured essentially the same trait as the full 1992 data, and would serve to yield an accurate measure of the short-term trend from 1992 to 1994. In addition, numerous comparisons were made between the differences in the average item scores for 1992 and 1994 and those obtained in the NAEP scale. These comparisons uniformly yielded similar patterns of results, further supporting the adequacy of the trend measurement.

Figure 12-6

*Example Short-Term Trend Dichotomous Item (R013102, Age 13/Grade 8) Demonstrating Differential Item Functioning Across Assessment Years 1992 and 1994**

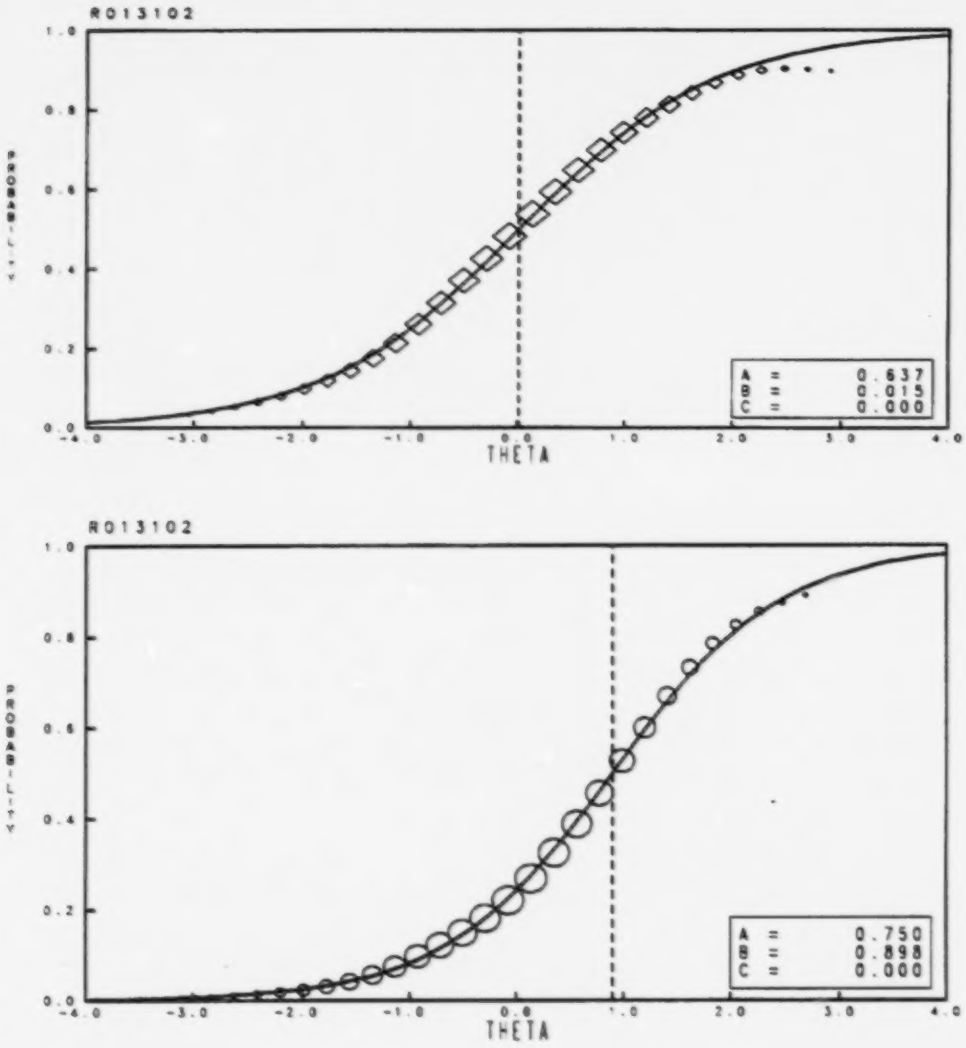


****Note****: The plot compares empirical and model-based estimates of the item response function (IRF). The smooth curve represents the model-based estimate at each provisional proficiency level. The diamonds represent 1994 age 13/grade 8 data; ovals represent 1992 age 13/grade 8 data.*

BEST COPY AVAILABLE

Figure 12-7

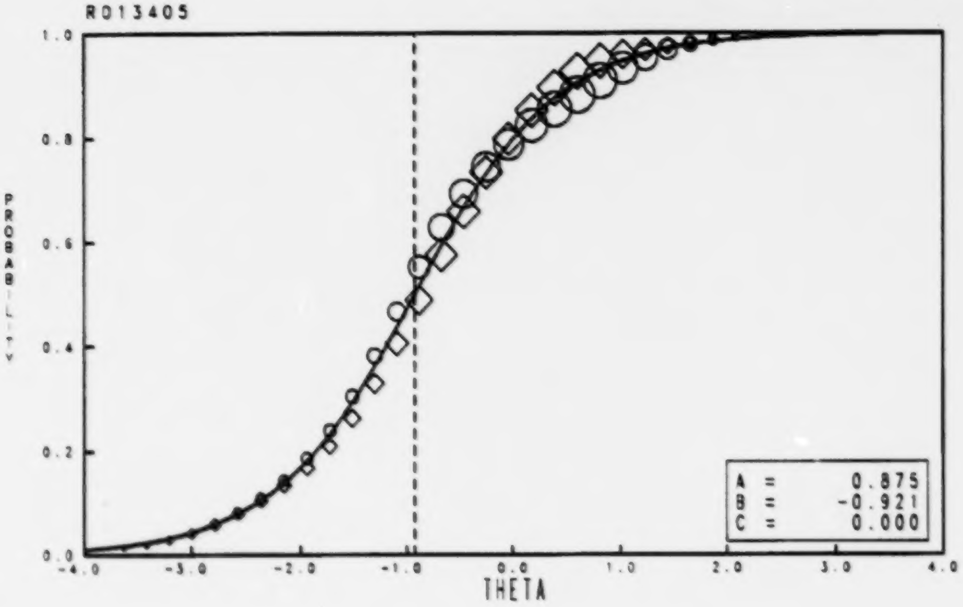
Example Short-Term Trend Dichotomous Item (R013102, Age 13/Grade 8)
*Fitting Separate Item Response Functions for Each Assessment Year**



****Note:*** *The plot compares empirical and model-based estimates of the item response function (IRF). The smooth curve represents the model-based estimate at each provisional proficiency level. The diamonds (top figure) represent 1994 age 13/grade 8 data; ovals (bottom figure) represent 1992 age 13/grade 8 data.*

Figure 12-8

Example Short-Term Trend Dichotomous Item (R013405, Age 17/Grade 12)
Demonstrating Acceptably Small Differential Item Functioning Across Assessment Years 1992 and



*1994**

Note: *The plot compares empirical and model-based estimates of the item response function (IRF). The smooth curve represents the model-based estimate at each provisional proficiency level. The diamonds represent 1994 age 17/grade 12 data; ovals represent 1992 age 17/grade 12 data.*

Table 12-13

Age 9/Grade 4 Items Scaled Separately by Assessment Year

Scale	Block	NAEP ID	Type
Reading for Literary Experience	R3	R012008	Short constructed-response
		R012010	Short constructed-response
	R4	R012102	Short constructed-response
		R012106	Short constructed-response
	R5	R012611	Short constructed-response
Reading to Gain Information	R6	R012201	Short constructed-response
		R012209	Multiple-choice
	R7	R012705	Short constructed-response
		R012706	Short constructed-response
		R012708	Extended constructed-response
		R012710	Short constructed-response
	R10	R012503	Short constructed-response
	R10	R012512	Extended constructed-response

Table 12-14
Age 13/Grade 8 Items Scaled Separately by Assessment Year

Scale	Block	NAEP ID	Type
Reading for Literary Experience	R3	R013102	Short constructed-response
	R5	R012607	Extended constructed-response
		R012611	Short constructed-response
Reading to Gain Information	R6	R013205	Short constructed-response
		R013211	Short constructed-response
	R7	R012705	Short constructed-response
		R012706	Short constructed-response
		R012708	Extended constructed-response
		R012710	Short constructed-response
Reading to Perform a Task	R10	R013402	Short constructed-response
		R013411	Short constructed-response
	R11	R013004	Extended constructed-response
		R013009	Short constructed-response
		R013011	Short constructed-response

Table 12-15

Age 17/Grade 12 Items Scaled Separately by Assessment Year

Scale	Block	NAEP ID	Type
Reading for Literary Experience	R3	R013102	Short constructed-response
	R4	R013503	Short constructed-response
		R013505	Short constructed-response
		R013508	Short constructed-response
		R013509	Short constructed-response
Reading to Gain Information	R6	R013201	Extended constructed-response
		R013205	Short constructed-response
		R013209	Short constructed-response
		R013211	Short constructed-response
		R013212	Extended constructed-response
	R7	R013701	Short constructed-response
		R013702	Short constructed-response
		R013704	Short constructed-response
		R013706	Short constructed-response
		R013708	Short constructed-response
		R013710	Short constructed-response
		R013712	Short constructed-response
	R13	R015503	Short constructed-response
		R015505	Short constructed-response
		R015509	Short constructed-response
		R015512	Short constructed-response
		R015514	Extended constructed-response
Reading to Perform a Task	R10	R013402	Short constructed-response
		R013411	Short constructed-response
		R013412	Short constructed-response
	R11	R013908	Short constructed-response
		R013910	Short constructed-response
		R013913	Short constructed-response

A list of the items scaled for each of the age/grades, along with their final item parameter estimates, appears in Appendix D.

12.5 GENERATION OF PLAUSIBLE VALUES

Multivariate plausible values were generated for each age/grade group separately using the CGROUP program described in Chapter 11. Final student weights were used in this analysis. Plans for reporting required analyses examining the relationships between proficiencies and a large number of background variables. The background variables included student demographic characteristics (i.e., race/ethnicity of the student, highest level of education attained by parents), students' perceptions about reading, student behavior both in and out of school (i.e., amount of television watched daily, amount of homework done each day), and a variety of other aspects of the educational, social, and financial environment of the schools they attended. For age 9/grade 4, information was also collected from students' teachers about the types of educational practice to which the students were exposed.

As described in Chapter 11, to avoid bias in reporting results and to minimize biases in secondary analyses, it was desirable to incorporate a large number of independent variables in the conditioning model. When expressed in terms of contrast-coded main effects and interactions, the number of variables to be included totaled 484 for age 9/grade 4, 533 for age 13/grade 8, and 467 for age 17/grade 12. The much larger numbers for age 9/grade 4 and age 13/grade 8 reflect the number of contrasts from the teachers' questionnaire. Appendix C provides a list of the full set of contrasts defined.

Some of these contrasts involved relatively small numbers of individuals and some were highly correlated with other contrasts or sets of contrasts. Given the large number of contrasts, an effort was made to reduce the dimensionality of the predictor variables. Consistent with what was done for the 1992 reading assessment, the original background variable contrasts were standardized and transformed into a set of linearly independent variables by extracting separate sets of principal components at each age/grade level. The principal components, rather than the original variables, were used as the independent variables in the conditioning model. The number of principal components was the number required to account for at least 90 percent of the variance in the original contrast variables. Research based on data from the 1990 Trial State Assessment in mathematics suggests that results obtained using such a subset of components will differ only slightly from those obtained using the full set (Mazzeo, Johnson, Bowker, & Fong, 1994). Table 12-16 contains a list of the number of principal components included in conditioning, as well as the proportion of variance accounted for by the conditioning model for each age/grade.

For each age/grade, Table 12-17 provides an estimated residual variance for each purpose-for-reading scale and the residual correlation matrix between the reading scales. The values, taken directly from the output of the CGROUP program, are estimates of relationships between the subscales conditional on the set of principal components included in the conditioning model. The marginal correlations between the purpose-for-reading scales are presented in Table 12-18. As expected, they are higher than the conditional correlations.

Table 12-16

Proportion of Proficiency Variance Accounted for by the Conditioning Model for the Reading Main Assessment

Grade	Number of Conditioning Contrasts*	Number of Principal Components*	Proportion of Proficiency Variance		
			Reading for Literary Experience	Reading to Gain Information	Reading to Perform a Task
4	484	218	.608	.631	—
8	533	238	.560	.607	.629
12	467	166	.626	.600	.652

*Note: *Excluding the constant term.*

Table 12-17

Conditional Correlations and Variances from Conditioning (CGROUP)

Grade	Scale	Reading for Literary Experience	Reading to Gain Information	Reading to Perform a Task
4	Reading for Literary Experience	1.000	–	–
	Reading to Gain Information	0.821	1.000	–
	Residual Variance	0.372	0.372	–
8	Reading for Literary Experience	1.000	–	–
	Reading to Gain Information	0.727	1.000	–
	Reading to Perform a Task	0.726	0.765	1.000
	Residual Variance	0.455	0.373	0.340
12	Reading for Literary Experience	1.000	–	–
	Reading to Gain Information	0.753	1.000	–
	Reading to Perform a Task	0.601	0.623	1.000
	Residual Variance	0.403	0.403	0.397

Table 12-18

*Marginal Correlations of Reading Scales**

Grade	Scale	Reading for Literary Experience	Reading to Gain Information	Reading to Perform a Task
4	Reading for Literary Experience	1.000	–	–
	Reading to Gain Information	0.881	1.000	–
8	Reading for Literary Experience	1.000	–	–
	Reading to Gain Information	0.829	1.000	–
	Reading to Perform a Task	0.822	0.858	1.000
12	Reading for Literary Experience	1.000	–	–
	Reading to Gain Information	0.859	1.000	–
	Reading to Perform a Task	0.782	0.809	1.000

****Note***: *These correlations are based on the "grade-only" portion of each sample. Tabled values were obtained by computing a separate Pearson correlation coefficient for each plausible value, computing Fisher's z-transformation for each value, computing the average of the transformed values, and computing the inverse transformation of the average.*

12.6 THE FINAL READING SCALES

12.6.1 Purpose-for-Reading Scales

The linear indeterminacy of the short-term trend scale was resolved by linking the 1994 short-term trend scales to previous scales. For each grade, the item parameters from the joint calibration based on data from 1992 and 1994 were used with the 1992 data to find plausible values for the 1992 data. The mean and standard deviation of all of the plausible values were calculated and matched to the mean and standard deviation of all of the plausible values based on the original analysis of the 1992 data, as given in earlier reports. This linking was performed separately for each of the purpose-for-reading scales using the transformation:

$$\theta_{\text{proficiency}} = A \bullet \theta_{\text{calibrated}} + B.$$

where $\theta_{\text{proficiency}}$ denotes values on the final transformed scale and $\theta_{\text{calibrated}}$ denotes values on the original calibration scale from BILOG/PARSCALE. The constants for the linear transformation for each scale are given in Table 12-19.

BEST COPY AVAILABLE

Table 12-19

Coefficients of Linear Transformations of the Purpose-for-Reading Scales from the Calibrating Scale Units to the Units of Reporting Proficiency

Age/Grade	Scale	A	B
9/4	Reading for Literary Experience	44.72	209.41
	Reading to Gain Information	44.78	205.39
13/8	Reading for Literary Experience	38.47	253.74
	Reading to Gain Information	40.14	254.37
	Reading to Perform a Task	45.15	253.36
17/12	Reading for Literary Experience	45.65	281.12
	Reading to Gain Information	33.97	286.57
	Reading to Perform a Task	39.61	283.13

12.6.2 The Composite Reading Scale

For the national assessment, a composite scale was created as an overall measure of reading proficiency. The composite was a weighted average of plausible values on the purpose-for-reading scales (Reading for Literary Experience, Reading to Gain Information, and, at age 13/grade 8 and age 17/grade 12, Reading to Perform a Task). The weights for the scales were proportional to the importance assigned to each reading purpose contained in the assessment specifications given in the *Reading Framework*. The weights are given in Table 12-20. As indicated in Chapter 2, the weights for each reading purpose are similar to the actual proportion of assessment time devoted to that purpose. In developing the composite scale, the weights were applied to the plausible values for each reading purpose as expressed in terms of the final scale (i.e., after transformation from the provisional BILOG/PARSCALE scales).

Table 12-20

Weights Used for Each Purpose-for-Reading Scale to Form the Reading Composite

Scale	Age 9/Grade 4	Age 13/Grade 8	Age 17/Grade 12
Reading for Literary Experience	.55	.40	.35
Reading to Gain Information	.45	.40	.45
Reading to Perform a Task	—	.20	.20

Overall summary statistics for the composite scale are given in Table 12-21.

Table 12-21

*Means and Standard Deviations on the Reading Composite Scale**

Grade	Year	Mean	S. D.
4	1994 1992	214.26 216.74	40.58 35.57
8	1994 1992	259.64 260.04	36.75 35.89
12	1994 1992	287.35 292.15	36.66 32.81

**Note: Tabled values were computed separately for each plausible value. The mean is the mean of the individual means. The standard deviation is computed as the square root of the average of the individual variances.*

12.7 PARTITIONING OF THE ESTIMATION ERROR VARIANCE

For each grade, the variance of the final, transformed scale mean was partitioned as described in Chapter 11. This analysis yielded estimates of the proportion of error variance due to sampling students and the proportion due to the latent nature of θ. These estimates are given in Table 12-22 for each purpose-for-reading scale and the composite scale (for stability of the estimates, they are based on 100 plausible values). Additional results, including those by gender and race/ethnicity, are presented in Appendix E.

Table 12-22

Estimation Error Variance and Related Coefficients for the Reading Main Assessment

Grade	Scale	Total Estimation Error Variance	Proportion of Variance Due to ...	
			Student Sampling	Latency of θ
4	Reading for Literary Experience	1.10	0.91	0.09
	Reading to Gain Information	0.97	0.90	0.10
	Composite	0.90	0.94	0.06
8	Reading for Literary Experience	0.96	0.84	0.16
	Reading to Gain Information	0.75	0.88	0.12
	Reading to Perform a Task	0.94	0.83	0.17
	Composite	0.70	0.93	0.07
12	Reading for Literary Experience	0.86	0.67	0.33
	Reading to Gain Information	0.34	0.76	0.24
	Reading to Perform a Task	0.70	0.71	0.29
	Composite	0.42	0.86	0.14

12.8 READING TEACHER QUESTIONNAIRE

Teachers of fourth- and eighth-grade students were surveyed about their educational background and teaching practices. This was done for the grade-eligible sample; teacher data was not collected for age-only eligible students. The students in the fourth grade were matched with their reading teacher. Eighth-grade students were matched first with a reading teacher, and then the specific classroom period. Variables derived from the questionnaire were used in the conditioning models. An additional conditioning variable was included that indicated whether the student had been matched with a teacher record. This contrast controlled estimates of subgroup means for differences that exist between matched and non-matched students. Of the 7,382 fourth-grade students in the sample, 6,793 (92%, unweighted) were matched with teachers who answered both parts of the teacher questionnaire, and 148 (2%, unweighted) of the students had teachers who answered only the teacher background section of the questionnaire. For the eighth-grade sample, 12,116 of the 15,606 students (78%, unweighted) were matched to both sections of the teacher questionnaire. An additional 1,655 students (11%, unweighted) were matched with the first part of the teacher questionnaire, but could not be matched to the appropriate classroom period. Thus, 94 percent of the fourth graders and 88 percent of the eighth graders were matched with at least the background information about their reading teacher.

Chapter 13

DATA ANALYSIS FOR THE U.S. HISTORY ASSESSMENT[1]

Bruce A. Kaplan, Eiji Muraki, John J. Ferris, and Steve Wang
Educational Testing Service

13.1 INTRODUCTION

This chapter describes the analyses performed on the responses to the cognitive and background items in the 1994 assessment of U.S. history. These analyses led to the results presented in *1994 NAEP U.S. History: A First Look—Findings from the National Assessment of Educational Progress* (Williams, Lazer, Reese, & Carr, 1995), and *NAEP 1994 U.S. History Report Card: Findings from the National Assessment of Educational Progress* (Beatty, Reese, Persky, & Carr, 1996). The emphasis of this chapter is on the methods and results of procedures used to develop the IRT-based scale scores that formed the basis of these reports. However, some attention is given to the analysis of constructed-response items as reported in the *NAEP 1994 U.S. History Report Card: Findings from the National Assessment of Educational Progress*. The theoretical underpinnings of the IRT and plausible values methodology described in this chapter are given in Chapter 11, and several of the statistics are described in Chapter 9.

The objective of the U.S. history analyses was to prepare scale values for the analysis of the focused-BIB U.S. history samples. The major analysis components are discussed in turn. Some aspects of the analysis, such as procedures for item analysis, scoring of constructed-response items, and methods of scaling, are described in previous chapters and are therefore not detailed here.

The analysis procedures carried out in the development of the 1994 U.S. history scales are based on the philosophical and theoretical underpinnings described in Chapter 11. There were seven major steps in the analysis of the 1994 U.S. history assessment, each of which is described in a separate section:

- conventional item and test analyses (Section 13.3.1);
- differential item functioning (DIF) analysis (Section 13.3.3);
- item response theory (IRT) within-grade scaling (Section 13.4.1);
- linking of the within-grade scales (Section 13.4.3);
- estimation of national and subgroup proficiency distributions based on the "plausible values" methodology (Section 13.5);
- transformation of the themes-of-history scales to the proficiency metric (Section 13.6.1); and
- creation of the composite U.S. history scale (Section 13.6.2).

---

[1] Bruce A. Kaplan specified most of the U.S. history analyses, coordinated the analyses, and completed most of the post-scaling analyses. Eiji Muraki advised and reviewed the scaling of the U.S. history data. Steve Wang and John J. Ferris carried out most of the scaling analyses and some of the post-scaling analyses. A portion of the analyses for this chapter were also performed by Norma A. Norris and Phillip T. Leung. James E. Carlson, Eugene Johnson, Edward Kulick, and Robert Mislevy consulted on DIF, IRT scaling, and generation of plausible values. John R. Donoghue gave direction and organization for this chapter.

BEST COPY AVAILABLE

13.2 DESCRIPTION OF STUDENT SAMPLE, ITEMS AND ASSESSMENT BOOKLETS

The 1994 U.S. history assessment will be used to start a new baseline for measuring trends in the nation. U.S. history was assessed by NAEP in 1986 and 1988. The 1994 assessment was a completely new assessment, designed with a new framework. Hence, there was no attempt to link it to previous assessments.

The student samples that were administered U.S. history items in the 1994 assessment are shown in Table 13-1. (See Chapters 1 and 3 for descriptions of the target populations and the sample design used for the assessment.) Data from the three samples (denoted His-MainP) were used in the analysis.

Table 13-1
NAEP 1994 U.S. History Student Samples

Sample	Booklets	Mode	Cohort Assessed	Time of Testing	Age Definition	Modal Grade	Number Assessed
9 [His-MainP]	101-118	Print	Age 9/ Grade 4	1/3/94-4/1/94	CY	4	7,205
13 [His-MainP]	101-133	Print	Age 13/ Grade 8	1/3/94-4/1/94	CY	8	11,669
17 [His-MainP]	101-133	Print	Age 17/ Grade 12	1/3/94-4/1/94	CY	12	10,837

LEGEND

His	*U.S. history*
MainP	*Main assessment, print administration*
Print	*Print administration*
CY	*Calendar year: birthdates in 1984, 1980, and 1976, respectively for ages 9, 13, and 17*

The items in the assessment were based on the curriculum framework described in *U.S. History Framework for the 1994 National Assessment of Educational Progress* (National Assessment Governing Board, 1994). Compared to previous NAEP assessments, the 1994 U.S. history assessment is "...far richer, requiring broad knowledge and analytical skills" (NAGB, 1994). As described in the U.S. history framework, these items are organized into four scales, corresponding to four themes of U.S. history:

1. *Democracy* - Change and Continuity in American Democracy: Ideas, Institutions, Practices, and Controversies
2. *Cultures* - The Gathering and Interactions of Peoples, Cultures, and Ideas
3. *Technology* - Economic and Technological Changes and Their Relation to Society, Ideas, and the Environment
4. *World Role* - The Changing Role of America in the World

In addition, a composite scale for U.S. history was created as a weighted sum of the themes-of-history scales.

The data from the focused-BIB assessment of U.S. history (9[His-MainP], 13[His-MainP], and 17[His-MainP]) were used for analyses comparing the levels of U.S. history achievement for various subgroups of the 1994 target populations. The assessment included three student cohorts: students who were either in the fourth grade or 9 years old, students who were either in the eighth grade or 13 years old, and students who were either in the twelfth grade or 17 years old. The birthdate ranges for age-eligible students were based on the 1984, 1980, and 1976 calendar years respectively for ages 9, 13, and 17. The sampled students in each of these three cohorts were assessed in the winter. The samples in the assessment are listed in Table 13-1.

In the samples, each student was administered a booklet containing a block of background questions common to all booklets for a particular age/grade level, two separately timed blocks of cognitive U.S. history items, a block of U.S. history-related background questions common to all U.S. history booklets for a particular age/grade level, and a block concerning the student's motivation and perception of the difficulty of the cognitive items. Six 25-minute blocks of U.S. history items were administered to the age 9/grade 4 students, and eight 25-minute blocks of U.S. history items were administered to the age 13/grade 8 and age 17/grade 12 students. As described in Chapter 2, the 25-minute blocks were combined into booklets according to a partially balanced incomplete block (BIB) design. (See Chapter 4 for more information about the blocks and booklets.) In addition, 50-minute U.S. history blocks were presented to the older students, one at grade 8 and one at grade 12. The 25-minute blocks were composed of mixtures of items from the four themes. The 50-minute block items were essentially one-theme blocks (Cultures in grade 8 and Technology in grade 12). The 50-minute blocks were included with the shorter blocks in the scaling. Both age- and grade-selected students contributed to the scaling. However, only the "grade-only" portion of the focused-BIB U.S. history samples contributed to the means and percentages of the results that are reported in the *NAEP 1994 U.S. History Report Card: Findings from the National Assessment of Educational Progress.*

Each block consisted of multiple-choice and constructed-response items. Constructed-response items were scored by specially trained readers (described in Chapter 7). Some of the constructed-response items were "timeline" items that required a few marks on a timeline and only a few sentences. These short constructed-response items were scored dichotomously as correct or incorrect. In addition, each block contained at least one extended constructed-response item, or a regular constructed-response item that required a more in-depth, elaborated response. These extended constructed-response items were scored polytomously on a five-point (0-4) scale, while the regular constructed-response items were scored polytomously on a four-point (0-3) scale. During the scaling process, the 0 category (omitted and off-task responses) was recoded. Off-task and omitted responses were combined with the adjacent category, "Unsatisfactory." This resulted in a 0-to-3 scale for the extended constructed-response items and a 0-to-2 scale for the regular constructed-response items. The extended constructed-response scale was

0 = Unsatisfactory (including off-task and omitted)
1 = Partial
2 = Essential
3 = Extensive, which demonstrates more in-depth understanding

The regular constructed-response scale was

0 = Unsatisfactory (including off-task and omitted)
1 = Partial
2 = Appropriate

In addition, categories of a small number of items were combined (collapsed). These changes were made so that the scaling model used for these items fit the data more closely, and are described more fully in Section 13.4.

The format and content of each block of items is given in Table 13-2. Common labeling of these blocks across age/grade levels does not denote common items. The numbers of items scaled in 1994 for each grade are presented in Table 13-3.

13.3 ITEM ANALYSES

13.3.1 Conventional Item and Test Analyses

Tables 13-4, 13-5, and 13-6 show the number of items in the block, the average weighted item score, average weighted polyserial correlation, and the weighted alpha reliability for each block administered. These statistics are described in Chapter 9. These values were calculated for the items within each block used in the scaling process. The tables also give the number of students who were administered the block and the percent attempting the last item in the block. These numbers include only those students in the grade-only portion of the samples that contributed to the summary statistics provided in the *NAEP 1994 U.S. History Report Card: Findings from the National Assessment of Educational Progress.*

Student weights were used for all statistics except sample sizes. The results for the blocks administered to each age/grade level indicate that the blocks differ in number of items, average difficulty, reliability, and percent attempting the last item, and so are not parallel to one another. Preliminary item analyses for all items within a block were completed before scaling; however, the results shown here indicate the characteristics of the items that contributed to the final scale.

As described in Chapter 9, in NAEP analyses (both conventional and IRT-based), a distinction is made between missing responses at the end of each block (not-reached) and missing responses prior to the last observed response (omitted). Items that were not reached were treated as if they had not been presented to the examinee, while omitted items were regarded as incorrect. The proportion of students attempting the last item of a block (or, equivalently, one minus the proportion not reaching the last item) is often used as an index of the degree of speededness of the block of items.

Standard practice at ETS is to treat all nonrespondents to the last item as if they had not reached the item. For multiple-choice items, short constructed-response items, and regular constructed-response items, this convention produced a reasonable pattern of results in that the proportion reaching the last item does not differ markedly from the proportion attempting the next-to-last item. However, for the blocks that ended with extended constructed-response items, this convention resulted in an implausibly large drop in the number of students attempting the final item. Therefore, for blocks that ended with an extended constructed-response item, students who attempted the next-to-last item but did not respond to the last item were classified as having intentionally omitted that item.

Table 13-2

Item Type Composition of Blocks Used in the U.S. History Assessment

Grade	Type of Item	Block								
		H3	H4	H5	H6	H7	H8	H9	H10	H11*
4	Number of multiple-choice items	10	9	11	10	11	11	–	–	–
	Number of short constructed-response items	0	0	0	0	0	1	–	–	–
	Number of regular constructed-response items	4	4	4	4	4	4	–	–	–
	Number of extended constructed-response items	1	2	1	1	1	1	–	–	–
	Total	15	15	16	15	16	17	–	–	–
8	Number of multiple-choice items	12	12	12	12	11	11	13	12	6
	Number of short constructed-response items	0	0	0	0	0	1	0	0	0
	Number of regular constructed-response items	4	4	4	4	4	4	3	4	5
	Number of extended constructed-response items	1	1	1	1	1	1	1	1	2
	Total	17	17	17	17	16	17	17	17	13
12	Number of multiple-choice items	12	13	10	12	12	12	12	12	8
	Number of short constructed-response items	0	1	0	0	0	0	0	0	0
	Number of regular constructed-response items	3	2	5	3	3	4	3	4	6
	Number of extended constructed-response items	2	2	2	2	2	1	1	1	5
	Total	17	18	17	17	17	17	17	17	19

***Note**: *This block is the 50-minute block for grades 8 and 12.*

BEST COPY AVAILABLE

Table 13-3

Numbers of Scaled U.S. History Items by Grade Levels

Grade	Democracy	Cultures	Technology	World Role	Total
4 only	18	20	17	6	**61**
8 only	24	28	17	12	**81**
12 only	29	28	34	30	**121**
4 and 8 only	12	9	5	7	**33**
8 and 12 only	10	4	10	10	**34**
4 and 12 only	0	0	0	0	**0**
4, 8, and 12	0	0	0	0	**0**
Total	**93**	**89**	**83**	**65**	**331**

Table 13-4

Descriptive Statistics for Item Blocks for the U.S. History Sample, Grade 4

Statistic	Block					
	H3	H4	H5	H6	H7	H8
Number of scaled items	15	15	16	15	16	17
Unweighted sample size	1,846	1,806	1,868	1,866	1,750	1,837
Average weighted item score	.41	.45	.52	.46	.40	.38
Average weighted r-polyserial	.50	.55	.57	.54	.53	.47
Weighted alpha reliability	.69	.74	.78	.74	.72	.68
Weighted proportion of students attempting last item	.83	.74	.95	.72	.88	.84

Table 13-5

Descriptive Statistics for Item Blocks for the U.S. History Sample, Grade 8

Statistic	Block								
	H3	H4	H5	H6	H7	H8	H9	H10	H11
Number of scaled items	17	17	17	17	16	17	17	17	13
Unweighted sample size	1,736	1,738	1,778	1,753	1,664	1,718	1,751	1,771	1,775
Average weighted item score	.55	.43	.43	.47	.58	.64	.39	.39	.43
Average weighted r-polyserial	.58	.58	.58	.52	.55	.58	.48	.51	.62
Weighted alpha reliability	.75	.77	.76	.73	.74	.76	.63	.67	.78
Weighted proportion of students attempting last item	.97	.98	.90	.97	.99	.99	.93	.88	100.

Table 13-6

Descriptive Statistics for Item Blocks for the U.S. History Sample, Grade 12

Statistic	Block								
	H3	H4	H5	H6	H7	H8	H9	H10	H11
Number of scaled items	17	18	17	17	17	17	17	17	19
Unweighted sample size	1,533	1,541	1,576	1,542	1,543	1,557	1,566	1,521	1,576
Average weighted item score	.43	.42	.48	.42	.42	.44	.52	.50	.46
Average weighted r-polyserial	.56	.58	.57	.57	.57	.55	.58	.55	.56
Weighted alpha reliability	.76	.79	.79	.75	.76	.75	.76	.74	.79
Weighted proportion of students attempting last item	.94	.89	.93	.97	.97	.97	.91	.91	.83

The results in Tables 13-4 to 13-6 indicate that the difficulty and internal consistency of the blocks, as indicated by the alpha reliability, varied. Such variability is expected, because the blocks were not constructed to be parallel. Based on the proportion of students attempting the last item, all of the blocks appear to be somewhat speeded. This effect is larger for grade 4 than for the other grades.

13.3.2 Scoring the Constructed-Response Items

As indicated earlier, the U.S. history assessment included constructed-response items. Responses to these items were included in the scaling process. In addition, detailed analyses of the ordinal responses to the constructed-response items were also conducted, and are summarized in the *NAEP 1994 U.S. History Report Card: Findings from the National Assessment of Educational Progress*. Chapter 7 provides the ranges for percent agreement between raters for the items as they were originally scored. The percent agreement for the raters and Cohen's (1968) Kappa are given in Appendix H.

13.3.3 Differential Item Functioning

Prior to scaling, DIF analyses were conducted on the U.S. history data for grades 4, 8, and 12. The purpose of these analyses was to identify items that were differentially difficult for various subgroups and to reexamine such items to determine their fairness and the appropriateness of including them in the scaling.

DIF analyses on multiple-choice items were based upon the modification of the Mantel-Haenszel procedure as adapted by Holland and Thayer (1988), and as described in Chapter 9. For constructed-response items, a generalized Mantel-Haenszel procedure was used. This procedure is due to Mantel and Haenszel (1959) and Somes (1986), and is also described in Chapter 9.

The "grade-only" portion of the focused-BIB was used for DIF analyses. Sample sizes were large enough to compare male and female students, White and Black students, and White and Hispanic students. Weights were rescaled separately for each comparison, as described in Chapter 9. DIF analyses were conducted separately by grade. A given item was subjected to at least three, and as many as six, separate DIF analyses. Tables 13-7 and 13-8 summarize the results of DIF analyses.

For multiple-choice items, the DIF index generated by the Mantel-Haenszel procedure is commonly used by ETS to place items into one of three categories: "A," "B," or "C." "A" items exhibit little or no DIF, while "C" items exhibit a strong indication of DIF and should be examined more closely. The DIF index used for constructed-response items used categories "AA" and "CC," where "AA" is similar to "A" of the multiple-choice items and "CC" is similar to the "C" items of the multiple-choice scale. Positive values of the index indicate items that are differentially easier for the "focal" group (female, Black, or Hispanic students) than for the "reference" groups (male or White students). Similarly, negative values indicate items that are differentially harder for the focal group than the reference group. An item that was classified as a "C" or "CC" item in *any* analysis was identified for further examination. Thirteen "C" items and 14 "CC" items were identified in the U.S. history assessment.

Following standard practice at ETS for DIF analyses conducted on final forms, all "C" and "CC" items were reviewed by a committee of trained test developers and subject-matter specialists. As described in Chapter 9, such committees are charged with making judgments regarding whether the differential difficulty of an item is *unfairly* related to group membership. The committee assembled to review NAEP items included ETS staff and outside members with expertise in the field. The committee carefully examined each "C" and "CC" item to determine if either the language or contents would tend to make the item more difficult for an identified group of examinees. It was the committee's judgment that none of the "C" and "CC" items were functioning differentially due to factors irrelevant to test objectives. Hence, none of the items were removed from scaling due to DIF.

Table 13-7

U.S. History DIF Categories by Grade for Dichotomous Items*

Grade	DIF Category	Analysis		
		Male/Female	White/Black	White/Hispanic
4	C-	0	1	0
	B-	3	4	5
	A-	31	24	30
	A+	26	23	24
	B+	3	8	4
	C+	0	3	0
8	C-	1	0	0
	B-	8	4	5
	A-	49	39	42
	A+	39	50	48
	B+	5	6	7
	C+	0	3	0
12	C-	3	1	0
	B-	8	7	2
	A-	41	37	45
	A+	50	53	52
	B+	3	6	6
	C+	0	1	0

*****Note:*** *"A" means little or no indication of DIF, "B" means a weak indication of DIF, and "C" means a strong indication of DIF. "+" means easier for focal group and "-" means harder for focal group.*

Table 13-8

U.S. History DIF Categories by Polytomous Items*

Grade	DIF Category	Analysis		
		Male/Female	White/Black	White/Hispanic
4	CC-	0	0	0
	AA-	10	20	15
	AA+	20	10	16
	CC+	1	1	0
8	CC-	1	0	1
	AA-	14	31	25
	AA+	30	14	20
	CC+	1	2	0
12	CC-	1	1	0
	AA-	16	24	30
	AA+	29	25	21
	CC+	5	1	0

****Note:*** *"AA" means little or no indication of DIF and "CC" means a strong indication of DIF. "+" means easier for focal group and "-" means harder for focal group.*

13.4 IRT SCALING

13.4.1 Item Parameter Estimation

Separate IRT-based scales were developed for each of the U.S. history themes identified in the U.S. history framework. As described in Chapter 11, multiple-choice items were fit using a 3PL model. Short constructed-response items were fit using a 2PL model. Regular and extended constructed-response items were fit using the generalized partial credit model.

For calibration, all items that were not reached were treated as if they had not been presented to the examinees. Responses to constructed-response items that were off-task were treated as if they had been omitted. The treatment of omitted responses differed according to the item type. Omitted responses to multiple-choice items were treated as fractionally correct to compensate for guessing. Omitted responses to short constructed-response items were treated as incorrect, and omitted responses to regular and extended constructed-response items were assigned to the lowest category.

Item parameters were estimated within each grade for each U.S. history theme. Thus, four scales, one for each theme, were created for each grade. Each within-grade calibration was performed using all the available examinees in the age/grade sample. Student sampling weights were used for the analysis.

For linking, sampling weights were restandardized, to a sum of 1,000, to ensure that each age/grade sample had a similar sum of weights and approximately equal influence in the calibration.

Item responses were calibrated using the BILOG/PARSCALE program. Starting values were computed from item statistics based on the entire data set. BILOG/PARSCALE calibrations were done in two stages. At stage one, the proficiency distribution of each age/grade was constrained to be normally distributed with fixed prior. The values of the item parameters from this normal solution were then used as starting values for a second stage estimation run in which the proficiency distribution was estimated concurrently with item parameters. Calibration was concluded when changes in item parameters became negligibly small.

A complexity introduced by the 50-minute blocks in U.S. history is that those blocks of items must be linked in some way to the shorter blocks. This is complicated by the fact that no students received the shorter blocks in addition to the 50-minute blocks. Because the samples of students receiving each booklet are representative of the population as a whole, it was assumed that the distribution of student proficiency was the same for the students receiving the 50-minute blocks as for the students receiving the booklets containing the shorter blocks.

13.4.2 Evaluation of Model Fit

During and subsequent to item parameter estimation, evaluations of the fit of the IRT models were carried out for each of the items. Model fit of dichotomous items was evaluated by examining plots of nonmodel-based estimates of the expected proportion correct (conditional on proficiency) versus the proportion correct predicted by the estimated item response function. For extended constructed-response items, similar plots were produced for each item category response function (see Chapter 9).

For the majority of the items, the model fit was good. In making decisions about excluding items from the final scales, a balance was sought between being too stringent, hence deleting too many items and possibly damaging the content representativeness of the pool of scaled items, and being too lenient, hence including items with model fit poor enough to endanger the types of model-based inferences made from NAEP results. Items that clearly did not fit the model were not included in the final scales; however, a certain degree of misfit was tolerated for a number of items included in the final scales.

A few polytomous items received special treatment in the scaling. Table 13-9 lists polytomous items that were recoded for scaling or dropped from the analyses. A list of the items scaled for each age/grade, along with their final item parameter estimates, appears in Appendix D.

After the item parameter estimation of IRT models, the fit of the models was evaluated for each of the items. Plots that are empirically based and theoretically computed were visually compared with the aids of chi-square item fit statistics. For most items, the model fit reasonably well. Therefore, all items were retained for subsequent scoring phase.

Item H053402 for grade 12 seems to be the worst fit among all dichotomous items. Figure 13-1 gives the plot of this item. In the figure, the X-axis depicts proficiency and the Y-axis depicts the probability of a correct response. The solid line is the theoretical item response curve and the symbols are nonmodel-based predicted proportion (Mislevy & Sheehan, 1987). Poor fit of this item is mainly caused by the differences on the upper range of the scale. Outside of this range, the empirical proportions generally followed the logistic curve. Figure 13-2 gives the best fitted dichotomous item (H03301, grade

8). The item is moderately difficult (P=+0.59) and discriminates well for a wide range of the proficiency scale. The item is moderately difficult (P+ = 0.59) and discriminates well for a wide range on the scale.

Table 13-9

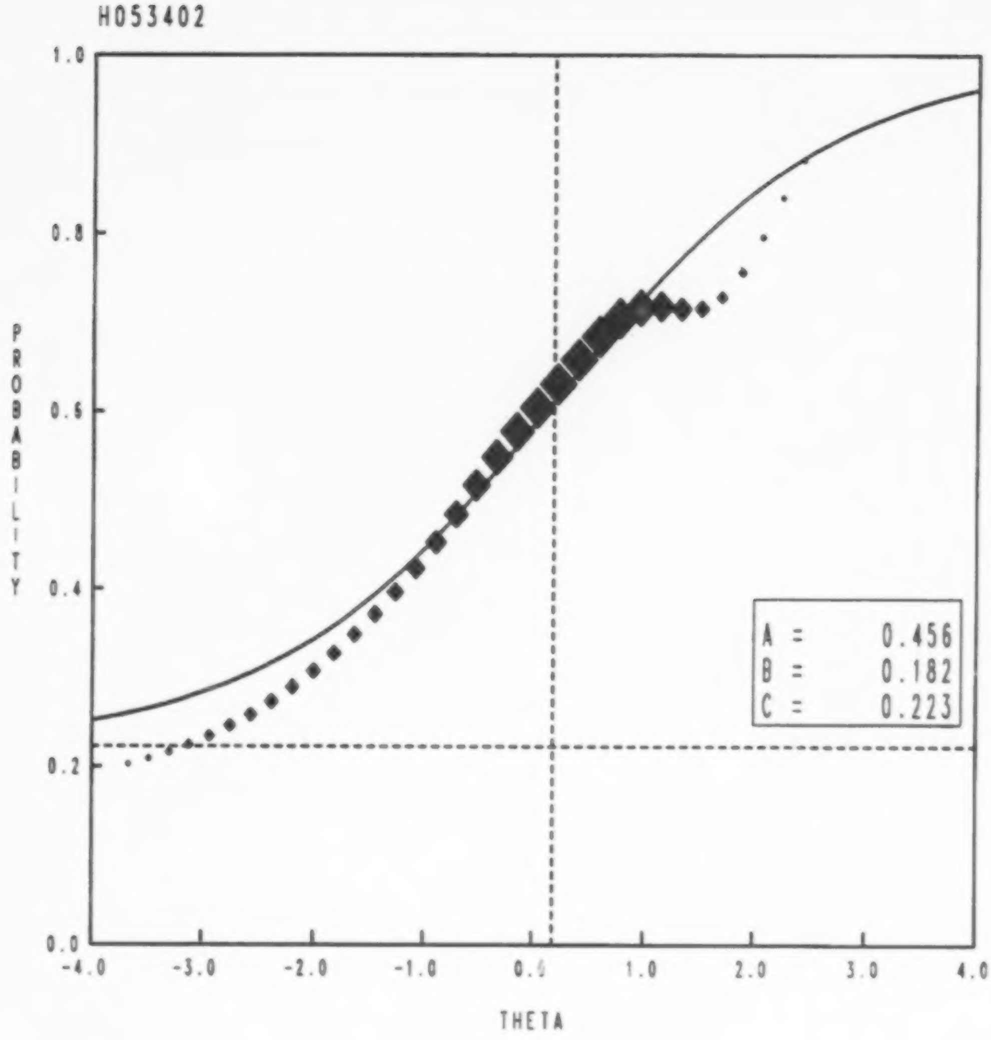
Items from the 1994 U.S. History Main Analysis Receiving Special Treatment

Scale	NAEP ID	Block	Grade(s) Affected	Disposition
Cultures	H030201	H4	4	Combine categories 3 and 4 (no one in category 4).
	H053801	H11	12	Item dropped (only item in 50-minute booklet in Cultures scale).
Technology	H053701	H11	12	Combine categories 3 and 4 (only 3 people in category 4).
World Role	H034101	H8	4	Combine categories 3 and 4 (only 3 people in category 4). No longer used as linking item to grade 8.

Figures 13-3 and 13-4 give the plots of each categorical response for polytomous extended constructed-response items. Item H028201 for grade 4 (see Figure 13-3) is an example of good fit and item H043501 for grade 8 was thought to be the worst fit among all polytomous items. As stated previously, in making decisions about excluding items from the final scales, a balance should be sought between being too stringent, hence deleting too many items and possibly damaging the content representativeness of the pool of scaled items, and being too lenient, hence including items with model fit poor enough to endanger the types of model-based inferences made from NAEP results. Item H043501 does not need to be dropped because the observed proportions generally follow the response curves the partial credit model predicts.

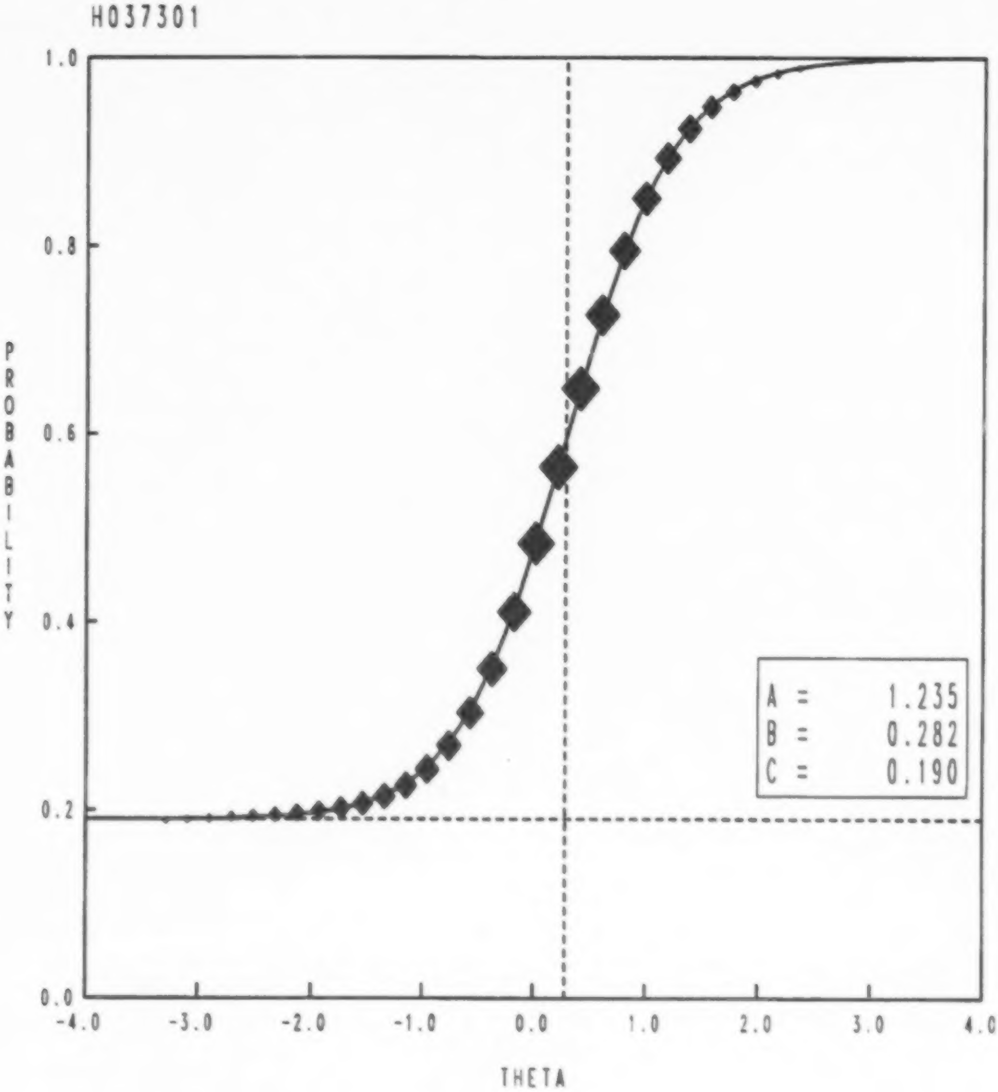
Figure 13-1

An Example of a Multiple-Choice (Dichotomous)
Item (H053402, Grade 12), Exhibiting Poor Model Fit



__Note__: The plot compares empirical and model-based estimates of the item response function (IRF). The smooth curve represents the model-based estimate at each provisional proficiency level. The diamonds represent the age 17/grade 12 data.

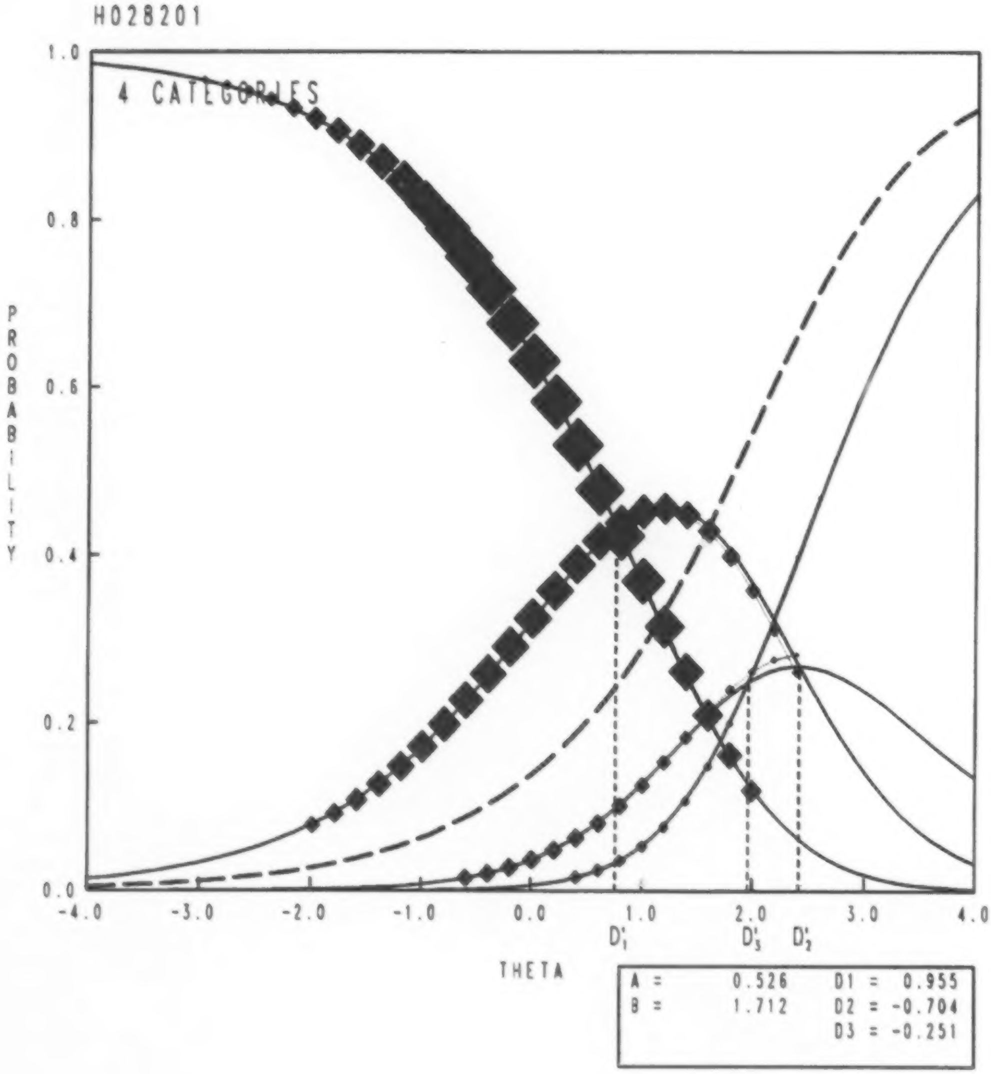
Figure 13-2
An Example of a Multiple-Choice (Dichotomous) Item (H037301, Grade 8) Exhibiting Good Model Fit



**__Note__: The plot compares empirical and model-based estimates of the item response function (IRF). The smooth curve represents the model-based estimate at each provisional proficiency level. The diamonds represent the age 13/grade 8 data.*

Figure 13-3

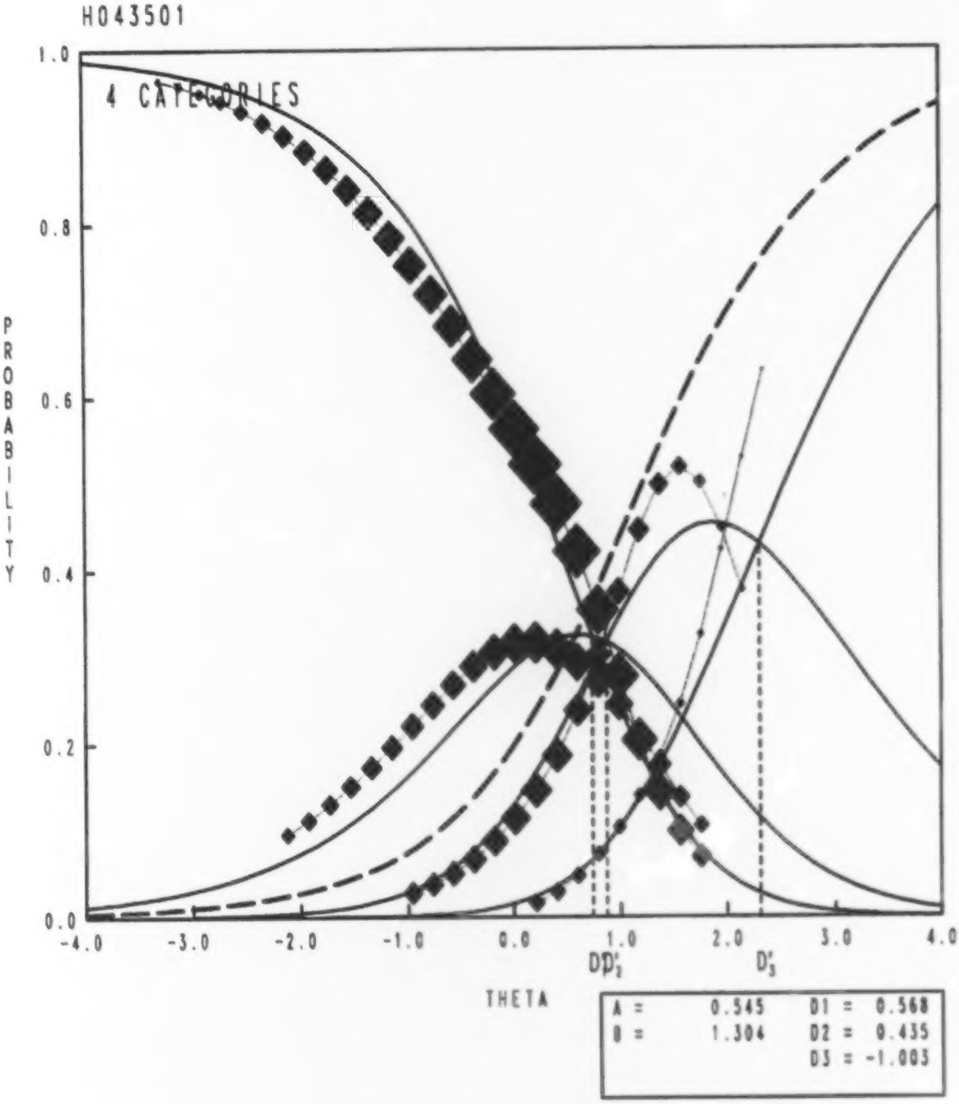
An Example of a Polytomous Item (H028201, Grade 4) Exhibiting Good Model Fit



****Note:*** *The plot compares empirical and model-based estimates of the item category response function (ICRF). The smooth curve represents the model-based estimate at each provisional proficiency level. The diamonds represent the age 9/grade 4 data.*

Figure 13-4

An Example of a Polytomous Item (H043501, Grade 8) Exhibiting Poor Model Fit



***Note:** *The plot compares empirical and model-based estimates of the item category response function (ICRF). The smooth curve represents the model-based estimate at each provisional proficiency level. The diamonds represent the age 13/grade 8 data.*

BEST COPY AVAILABLE

13.4.3 Linking of Within-Grade Scales

After the within-grade scaling described in Section 13.4.1 was completed for each of the four scales (themes) within each of the three grades, the weighted generalized Stocking-Lord (Muraki & Grima, 1993) procedure was used to place the data for the three grades on the same scale. For each scale, this was accomplished by linking the fourth-grade scale to the eighth-grade scale in one analysis, then linking the twelfth-grade scale to the eighth-grade scale in a second analysis. Using A and B constants produced by the generalized Stocking-Lord procedure, the transformed proficiency variable is

$$\theta^*_{ia} = A_a \theta_{ia} + B_a,$$

where θ and θ^* represent a proficiency variable before and after transformation, respectively, *i* indexes the five plausible values, and *a* indexes the grade (4, 8, or 12). The constants are displayed in Table 13-10. The constants for all eighth-grade scales were 1.0 and 0.0, since the other grades were linked to it.

Table 13-10
Linear Transformation Constants for the U.S. History Between Grade Linking

Grade	Scale	A_a	B_a
Grade 4 to Grade 8	Democracy	1.3769	-1.5130
	Cultures	1.7831	-1.9905
	Technology	1.4759	-1.6288
	World Role	0.8722	-1.4121
Grade 12 to Grade 8	Democracy	1.2726	0.9098
	Cultures	0.8860	0.6324
	Technology	0.8921	0.8056
	World Role	1.0538	0.7381

13.5 GENERATION OF PLAUSIBLE VALUES

Multivariate plausible values were generated for each age/grade group separately using the CGROUP program described in Chapter 11. Final student weights were used in this analysis. Plans for reporting necessitated analyses that examined the relationships between proficiencies and a large number of background variables. The background variables included student demographic characteristics (e.g., race/ethnicity of the student, highest level of education attained by parents), students' perceptions about U.S. history, student behavior both in and out of school (e.g., amount of television watched daily, amount of homework done each day), and a variety of other aspects of the educational, social, and financial environment of the schools they attended. For age 9/grade 4 and age 13/grade 8, information was also collected from students' teachers about the types of educational practice to which the students were exposed.

As described in Chapter 11, to avoid bias in reporting results and to minimize biases in secondary analyses, it was desirable to incorporate a large number of independent variables in the conditioning model. When expressed in terms of contrast-coded main effects and interactions, the number of variables totaled 589 for age 9/grade 4, 580 for age 13/grade 8, and 408 for age 17/grade 12. The much larger numbers for age 9/grade 4 and age 13/grade 8 reflect the inclusion of additional contrasts from the teacher questionnaires. Appendix C provides a list of the full set of contrasts.

Some of these contrasts involved relatively small numbers of individuals and some were highly correlated with other contrasts or sets of contrasts. Given the large number of contrasts, an effort was made to reduce the dimensionality of the predictor variables. In previous years, the original background variable contrasts were standardized and transformed into a set of linearly independent variables by extracting separate sets of principal components at each age/grade level. The principal components, rather than the original variables, were used as the independent variables in the conditioning model. The number of principal components was the number required to account for at least 90 percent of the variance in the original contrast variables. Research based on data from the 1990 Trial State Assessment in mathematics suggests that results obtained using such a subset of components will differ only slightly from those obtained using the full set (Mazzeo, Johnson, Bowker, & Fong, 1992). Table 13-11 contains a list of the number of principal components included in conditioning, as well as the proportion of variance accounted for by the conditioning model for each age/grade. This proportion is the ratio of the difference between the total variance and the CGROUP residual varaiance, divided by the total variance. The total variance is the mean of each of the five theta variances obtained by their respective plausible values.

For each age/grade, Table 13-12 provides estimated residual variances for each U.S. history theme scale and the residual correlation matrix between the U.S. history scales. The values, taken directly from the output of the CGROUP program, are estimates of relationships between the scales conditional on the set of principal components included in the conditioning model. The marginal correlations between the U.S. history theme scales are presented in Table 13-13. These marginal correlations were obtained by computing a separate Pearson correlation coefficient for each plausible value, computing Fisher's z-transformation for each correlation, then computing the average of the transformed values, and finally, computing the inverse Fisher z-transformation of the average. The marginal correlations are computed from the grade only sample.

Table 13-11

Proportion of Proficiency Variance Accounted for by the Conditioning Model for the U.S. History Assessment

Grade	Number of Conditioning Contrasts*	Number of Principal Components*	Proportion of Proficiency Variance			
			Democracy	Cultures	Technology	World Role
4	589	247	0.654	0.622	0.678	0.632
8	580	244	0.595	0.595	0.650	0.603
12	408	190	0.646	0.663	0.639	0.613

__Note__: Excluding the constant term.

Table 13-12

Conditional Correlations and Residual Variances from Conditioning (CGROUP)*

Age/Grade	Scale	Democracy	Cultures	Technology	World Role
9/4	Democracy	1.000	–	–	–
	Cultures	0.914	1.000	–	–
	Technology	0.875	0.941	1.000	–
	World Role	0.917	0.889	0.882	1.000
	Residual Variance	0.275	0.263	0.283	0.463
13/8	Democracy	1.000	–	–	–
	Cultures	0.902	1.000	–	–
	Technology	0.845	0.911	1.000	–
	World Role	0.877	0.820	0.775	1.000
	Residual Variance	0.437	0.394	0.281	0.407
17/12	Democracy	1.000	–	–	–
	Cultures	0.929	1.000	–	–
	Technology	0.940	0.906	1.000	–
	World Role	0.864	0.798	0.861	1.000
	Residual Variance	0.382	0.275	0.303	0.346

**Note*: *These correlations are based on the age and grade sample.*

Table 13-13
*Marginal Correlations of U.S. History Scales**

Grade	Scale	Democracy	Cultures	Technology	World Role
4	Democracy	1.000	–	–	–
	Cultures	0.900	1.000	–	–
	Technology	0.887	0.929	1.000	–
	World Role	0.860	0.874	0.870	1.000
8	Democracy	1.000	–	–	–
	Cultures	0.910	1.000	–	–
	Technology	0.880	0.911	1.000	–
	World Role	0.891	0.855	0.826	1.000
12	Democracy	1.000	–	–	–
	Cultures	0.944	1.000	–	–
	Technology	0.950	0.936	1.000	–
	World Role	0.888	0.847	0.883	1.000

***Note*: These correlations are based on the "grade only" portion of each sample. Tabled values were obtained by computing a separate Pearson correlation coefficient for each plausible value, computing Fisher's z-transformation for each value, computing the average of the transformed values, and computing the inverse transformation of the average.**

13.6 THE FINAL U.S. HISTORY SCALES

13.6.1 U.S. History Theme Scales

Like all IRT scales, the U.S. history scales have a linear indeterminacy that may be resolved by an arbitrary choice of origin and unit size for each scale. Similar to previous NAEP assessments, the scale mean and standard deviation were set to 250 and 50 using the transformation

$$\theta_{\text{proficiency}} = A \bullet \theta_{\text{calibrated}} + B,$$

where $\theta_{\text{proficiency}}$ denotes values on the final transformed scale and $\theta_{\text{calibrated}}$ denotes values on the original calibration scale from BILOG/PARSCALE.

For each of the four U.S. history theme scales, the linear indeterminacy was resolved by transforming the mean and standard deviation of the three age/grade samples combined together to the 250, 50 metric. The constants for the linear transformation, including the grade linking using the generalized Stocking-Lord procedure (see Section 13.4.3) for each scale, are given in Table 13-14.

Table 13-14

Coefficients of Linear Transformations of the U.S. History Themes Scales from the Calibrating Scale Units to the Units of the U.S. History Scale*

Age/Grade	Scale	A	B
9/4	Democracy	45.658	201.389
	Cultures	59.868	191.297
	Technology	53.389	194.277
	World Role	32.839	200.041
13/8	Democracy	33.160	251.561
	Cultures	33.575	258.129
	Technology	36.174	253.197
	World Role	37.653	253.211
17/12	Democracy	42.200	281.730
	Cultures	29.746	279.364
	Technology	32.271	282.333
	World Role	39.678	281.001

* *Note: This transformation incorporates the within-grade equating transformation discussed in Section 13.4.3.*

13.6.2 The Composite U.S. History Scale

For the national assessment, a composite scale was created as an overall measure of U.S. history proficiency. The composite was a weighted average of plausible values on the U.S. history theme scales. The weights for the scales were proportional to the importance assigned to each U.S. history theme contained in the assessment specifications given in the *U.S. History Framework*. The weights are given in Table 13-15. As indicated in Chapter 2, the weights for each of the U.S. history themes are similar to the actual proportion of assessment time devoted to that theme. In developing the composite scale, the weights were applied to the plausible values for each U.S. history theme as expressed in terms of the final scale (i.e., after transformation from the provisional BILOG/PARSCALE scales).

Table 13-15

Weights Used for Each U.S. History Theme Scale to Form the U.S. History Composite

Scale	Age 9/Grade 4	Age 13/Grade 8	Age 17/Grade 12
Democracy	.25	.30	.25
Cultures	.35	.30	.25
Technology	.25	.20	.25
World Role	.15	.20	.25

It is necessary to caution that, although the U.S. history composite is expressed in units that seem similar to other NAEP scales, it is not appropriate to compare scores to those of other subject areas. The transformation chosen to resolve the linear indeterminacy in the U.S. history composite is a convenient transformation, but it is only one of a conceptually infinite number of such transformations that could have been chosen. Any one of these transformations would have provided equivalent information about the relative standings of subgroups in the population. *Because there is no link, real or implied, in the construction of the U.S. history composite and the U.S. history theme scales to either the reading, geography, mathematics, science, or writing assessments, the comparison of students' U.S. history scores to students' scores in other subject areas is devoid of meaning.* Overall summary statistics for the composite scale are given in Table 13-16.

Table 13-16

Means and Standard Deviations on the U.S. History Composite Scale

Grade	All Five Plausible Values	
	Mean	S. D.
4	204.6	41.4
8	259.3	31.8
12	286.1	32.1

13.7 PARTITIONING OF THE ESTIMATION ERROR VARIANCE

For each grade, the variance of the final, transformed scale mean was partitioned as described in Chapter 11. This analysis yielded estimates of the proportion of error variance due to sampling students and the proportion due to the latent nature of θ. These estimates are given in Table 13-17 for each U.S. history theme scale and the composite scale (for stability of the estimates, they are based on 100 plausible values). Additional results, including those by gender and race/ethnicity, are presented in Appendix E.

371

Table 13-17

Estimation Error Variance and Related Coefficients for the U.S. History Assessment

Grade	Scale	Total Estimation Error Variance	Proportion of Variance Due to ...	
			Student Sampling	Latency of θ
4	Democracy	1.20	0.79	0.21
	Cultures	1.71	0.87	0.13
	Technology	1.56	0.88	0.12
	World Role	1.09	0.78	0.22
	Composite	1.09	0.95	0.05
8	Democracy	0.53	0.84	0.16
	Cultures	0.45	0.85	0.15
	Technology	0.49	0.75	0.25
	World Role	0.73	0.79	0.21
	Composite	0.38	0.94	0.06
12	Democracy	1.16	0.87	0.13
	Cultures	0.50	0.87	0.13
	Technology	0.46	0.86	0.14
	World Role	0.81	0.83	0.17
	Composite	0.60	0.95	0.05

13.8 U.S. HISTORY TEACHER QUESTIONNAIRE

Teachers of the fourth- and eighth-grade students were surveyed about their educational background and teaching practices. The students in the fourth grade were matched with their U.S. history teacher. Students in the eighth grade were matched first with a U.S. history teacher and then with the specific classroom period. This was done for a "grade-only" sample. Variables derived from the questionnaire were used in the conditioning models, along with a variable that indicated whether a student record had been matched with a teacher record, which controls estimates of subgroup means for differences that exist between matching and non-matching students. Of the 5,499 fourth-grade students in the sample, 88 percent were matched with teachers who answered both parts of the teacher questionnaire, and six percent of the students had teachers that only answered the teacher background section of the teacher questionnaire. For the eighth-grade sample, 77 percent were matched to both parts of the teacher questionnaire, and 10 percent were matched with only the first part (which does not determine the specific classroom period) of the teacher questionnaire. Thus, 94 percent of the fourth graders and 87 percent of the eighth graders were matched with at least the background information about their U.S. history teacher.

Chapter 14

DATA ANALYSIS FOR THE GEOGRAPHY ASSESSMENT[1]

Frank Jenkins and Edward Kulick
Educational Testing Service

14.1 INTRODUCTION

This chapter describes the analyses performed on the responses to the cognitive and background items in the 1994 assessment of geography. The focus of the chapter is on the methods and procedures used to estimate proficiency distributions for subgroups of students. The chapter describes a wide array of analysis topics, such as the scoring of constructed-response items, classical item characteristics, item response theory (IRT) analysis of the geography scale, and the imputation of plausible values of student proficiencies. The statistical basis of the IRT and plausible values methodology described in this chapter are given in Chapter 11. These analyses led to the results presented in *NAEP 1994 Geography: A First Look—Findings from the National Assessment of Educational Progress* (Williams, Reese, Lazer, & Shakrani, 1995) and *NAEP 1994 Geography Report Card: Findings from the National Assessment of Educational Progress* (Persky, Reese, O'Sullivan, Lazer, Moore, & Shakrani, 1996).

The main purpose of the geography analysis was to estimate subgroup proficiency distributions for national samples of students who were administered geography items according to a balanced incomplete block (BIB) design.

The student samples that were administered geography items in the 1994 assessment are shown in Table 14-1. Chapters 1 and 3 contain descriptions of the target populations and the sample design used for the assessment. Note that at grades 8 and 12, 19 booklets were used, one more than for grade 4. The 19th booklet contains a single 50-minute theme block and is not part of the BIB spiral design.

The 1994 geography samples were analyzed to provide comparisons of geography achievement for various subgroups of the 1994 target populations. The target populations were grade 4, grade 8, and grade 12 students in the U.S. Three student cohorts were used for estimating item parameters: students who were either in the fourth grade or 9 years old, students who were either in the eighth grade or 13 years old, and students who were either in the twelfth grade or 17 years old. The birthdate ranges for age-eligible students were based on the 1984, 1980, and 1976 calendar years respectively for ages 9, 13, and 17. The age/grade cohorts were used for item parameter estimation (scaling) purposes to assure an adequate number of students per item, but only the grade-defined subset of each cohort was used for reporting. The sampled students in each of these three cohorts were assessed in the winter (February to April).

---

[1] Frank Jenkins was the primary person responsible for the planning, specification, and coordination of the geography analyses. Computer activities for all geography scaling and data analyses were directed by Edward Kulick and completed by Y. Fai Fong and Steve Wang. Others contributing to the analysis of geography data were Bruce A. Kaplan, David S. Freund, Steven P. Isham, and Norma A. Norris.

Table 14-1

NAEP 1994 Geography Student Samples

Sample	Booklets*	Cohort Assessed	Time of Testing	Age Definition	Modal Grade	Number Assessed (Age/Grade)	Number Assessed (Grade)
9 [Geo-MainP]	31-48	Age 9/ Grade 4	1/3/94- 4/1/94	CY	4	7,205	5,507
13 [Geo-MainP]	31-49	Age 13/ Grade 8	1/3/94- 4/1/94	CY	8	9,186	6,878
17 [Geo-MainP]	31-49	Age 17/ Grade 12	1/3/94- 4/1/94	CY	12	8,589	6,234

**Note: Eighteen booklets were used for grade 4 and 19 booklets were used for grades 8 and 12.*

LEGEND

Geo	*Geography*
MainP	*Main assessment, print administration*
CY	*Calendar year: birthdates in 1984, 1980, and 1976, respectively for ages 9, 13, and 17*

14.2 DESCRIPTION OF ITEMS AND ASSESSMENT BOOKLETS

The pool of items used in the 1994 geography assessment contained a range of constructed-response and multiple-choice questions measuring performance on sets of objectives. The framework for the objectives is described in Chapter 2 and *NAEP 1994 Geography Report Card: Findings from the National Assessment of Educational Progress.* A total of 268 distinct geography items addressing these objectives were scaled after some items were dropped (see table 14-2). Note that some items overlap across grade. The total number of scaled items was 89, 124, and 119 respectively for grades 4, 8, and 12. The items were classified into three categories based on their content: Space and Place; Environment and Society; and Spatial Dynamics and Connections. Table 14-2 shows the numbers of items within content area scales for each grade. These three categories of items constituted the scales used in 1994 reporting. The items were partitioned into six blocks at all three grades and these blocks were then assigned to 18 booklets for each grade according to a balanced incomplete block (BIB) design. Tables 4-12 and 4-14 in Chapter 4 give the correspondence between booklets and blocks, and Tables 6-13, 6-15, and 6-16 in Chapter 4 give the correspondence between blocks and booklets and the number of types of items in blocks. Students were given 25 minutes to complete each of the six BIB blocks. At grades 8 and 12 there was a 19th booklet that consisted of a single block, which students were given 50 minutes to complete. Each booklet contained relatively few items from each of the three categories. The 19th booklet had no items in the Spatial Dynamics and Connections content area at grades 8 and 12.

Each student's assessment booklet contained one section of general background items, a second section of geography-specific background items, two sections of geography cognitive items (except for the 19th booklet, which had one cognitive section) and a section of motivation questions.

Both age-selected and grade-selected students contributed to the scaling (the estimation of item parameters). However, only the "grade-selected" portion of the geography samples contributed to the means and percentages in the results that are reported in the *NAEP 1994 Geography Report Card: Findings from the National Assessment of Educational Progress*. The 268 unique items were constructed according to several formats, some traditional and some more innovative: multiple-choice, short constructed-response, regular constructed-response, and extended constructed-response. The multiple-choice items conformed to the familiar format of a stem followed by several possible answers, with only one answer being correct. Short constructed-response items were questions that required an open-ended response (i.e., explaining why a particular answer had been given to a question). Responses to these questions were read by raters who assigned them a score of correct or incorrect. Regular and extended constructed-response items were open-ended questions that required several stages of reasoning or problem-solving. Responses to these types of items were also read by raters but were given a score reflecting the degree of correctness rather than simply being judged right or wrong. As Table 14-3 indicates, of the 89 items used at grade 4, 58 were multiple-choice and 31 (or 35 percent) were constructed-response of some sort. Of 124 items used at grade 8, 83 were multiple-choice and 41 (or 33 percent) were constructed-response. Of the 119 items scaled at grade 12, 81 were multiple-choice items and 38 (or 31 percent) were constructed-response. The count of items in the different scales is found in Table 14-4. Most items were in the Space and Place scale followed by Spatial Dynamics and Connections and finally by Environment and Society.

Table 14-2

Numbers of Scaled Geography Items
Common Across Grade Levels, by Content Area Scale

Grade	Space and Place	Environment and Society	Spatial Dynamics and Connections	Total Items	Total Items Scaled
4 only	23	21*	14	59	58
8 only	29	21	10	60	60
12 only	44*	18*	24*	89	86
4 and 8	14	4	13	31	31
8 and 12	15*	7	11	34	33
4 and 12	0	0	0	0	0
4, 8, and 12	0	0	0	0	0
Total	127	73	73	273	268

*Note: One item in this cell was deleted from scaling due to lack of fit with the IRT model. Numbers exclude deleted items.

Table 14-3

*Item Type Composition of Blocks Used in the Geography Assessment**

Grade	Type of Item	Block							
		G3	G4	GE	G6	G7	G8	G9	Total
4	Number of multiple-choice items	8	10	11	9	9	12	–	59
	Number of short constructed-response items	0	0	0	2	1	0	–	3
	Number of regular constructed-response items	4	3	3	4	3	4	–	21
	Number of extended constructed-response items	2	2	1	0	1	1	–	7
	Total	14	15	15	15	14	17	–	90
8	Number of multiple-choice items	12	12	12	11	9	12	16	84
	Number of short constructed-response items	0	1	0	0	1	0	0	2
	Number of regular constructed-response items	4	3	5	3	3	4	9	31
	Number of extended constructed-response items	1	1	1	2	1	1	1	8
	Total	17	17	18	16	14	17	26	125
12	Number of multiple-choice items	12	12	12	11	11	11	16	85
	Number of short constructed-response items	0	0	0	0	0	0	0	0
	Number of regular constructed-response items	2	2	5	3	3	4	5	24
	Number of extended constructed-response items	3	3	1	2	1	2	2	14
	Total	17	17	18	16	15	17	23	123

*Note: *Includes items deleted from scaling due to lack of fit.*

Table 14-4

Numbers of Items in Scales

Scales	Grade 4	Grade 8	Grade 12
Space and Place			
Scaled	37	58	59
Deleted	0	1	2
Total	37	59	61
Environment and Society			
Scaled	25	32	25
Deleted	1	0	1
Total	26	32	26
Spatial Dynamics and Connections			
Scaled	27	34	35
Deleted	0	0	1
Total	27	34	36
Total			
Scaled	89	124	119
Deleted	1	1	4
Total	90	125	123

14.3 ITEM ANALYSES

14.3.1 Conventional Item and Test Analyses

This section contains a detailed description of the conventional item analysis performed on the geography data. This analysis was done within block so that a student's score is the sum of item scores in a block. There were four types of items: multiple-choice and short constructed-response, which were scored dichotomously as right or wrong, and regular and extended constructed-response, which were given partial credit (polytomously scored). Polytomous items are not scored right/wrong but are scored with three or more categories reflecting several degrees of knowledge. Item statistics were calculated for mean percent correct, average score, item-to-total-score correlations (a biserial correlation for dichotomous items and a polyserial correlation for polytomous items), and percent responding in each item category.

Tables 14-5, 14-6, and 14-7 show the number of items, mean item score, mean item-to-total score correlation, and alpha reliability for each block administered at each age/grade level for the main assessment. These values were calculated within each block and only for those items used in the scaling process. The table also gives the number of students who were administered the block and the percent reaching the last item in the block. These numbers are based on the students in the grade-only portion of the samples that also formed the basis of the summary statistics provided in the *NAEP 1994 Geography Report Card: Findings from the National Assessment of Educational Progress.* Student weights were used, except for the sample sizes. The results for the blocks administered to each age/grade level indicate that the blocks differ in number of items, average difficulty, reliability, and percent not reaching the last item.

As described in Chapter 9, in NAEP analyses (both conventional and IRT-based), a distinction is made between missing responses at the end of each block (not reached) and missing responses prior to the last observed response (omitted). Items that were not reached were treated as if they had not been presented to the examinee, while omitted items were regarded as incorrect. The proportion of students attempting the last item of a block (or, equivalently, one minus the proportion not reaching the last item) can be used as an index of the degree of speededness of the block of items.

Standard practice at ETS is to treat all nonrespondents to the last item as if they had not reached the item. For multiple-choice items, short constructed-response items, and regular constructed-response items, this convention produced a reasonable pattern of results, in that the proportion reaching the last item does not differ markedly from the proportion attempting the next to last item. However, for the blocks that ended with extended constructed-response items, this convention resulted in an implausibly large drop in the number of students attempting the final item. Therefore, for blocks that ended with an extended constructed-response item, students who attempted the next-to-last item but did not respond to the last item were classified as having intentionally omitted that item.

Table 14-5

*Descriptive Statistics for Item Blocks from the Geography Main Sample, Grade 4**

	Block					
Statistic	**G3**	**G4**	**G5**	**G6**	**G7**	**G8**
Number of scaled items	14	15	14	15	14	17
Unweighted sample size	1,816	1,822	1,841	1,800	1,849	1,859
Average item score	.54	.55	.57	.48	.37	.44
Average weighted item-to-total correlation	.61	.60	.62	.58	.58	.53
Weighted alpha reliability	.79	.79	.77	.75	.74	.73
Weighted proportion of students reaching the last item	.79	.94	.90	.91	.61	.77

****Note****: For dichotomous items, average item scores are the simple percent correct and for polytomous items, these scores are item averages scaled to go from 0 to 1.*

Table 14-6

*Descriptive Statistics for Item Blocks from the Geography Main Sample, Grade 8**

Statistic	Block						
	G3	G4	G5	G6	G7	G8	G9
Number of scaled items	17	17	17	16	14	17	26
Unweighted sample size	1,700	1,734	1,706	1,702	1,705	1,697	1,726
Average item score	.50	.48	.50	.51	.64	.68	.48
Average weighted r-biserial	.59	.60	.62	.54	.68	.62	.54
Weighted alpha reliability	.77	.78	.80	.74	.82	.77	.82
Weighted proportion of students reaching the last item	.98	.98	.97	.87	.91	.93	.99

***Note**: *For dichotomous items, average item scores are the simple percent correct and for polytomous items, these scores are item averages scaled to go from 0 to 1.*

Table 14-7

*Descriptive Statistics for Item Blocks from the Geography Main Sample, Grade 12**

Statistic	Block						
	G3	G4	G5	G6	G7	G8	G9
Number of scaled items	16	16	17	16	15	17	22
Unweighted sample size	1,558	1,540	1,572	1,523	1,543	1,544	1,563
Average item score	.55	.54	.63	.63	.67	.42	.52
Average weighted r-biserial	.60	.57	.61	.59	.64	.58	.60
Weighted alpha reliability	.77	.77	.79	.77	.79	.78	.83
Weighted proportion of students reaching the last item	.93	.96	.98	.92	.58	.74	.99

***Note**: *For dichotomous items, average item scores are the simple percent correct and for polytomous items, these scores are item averages scaled to go from 0 to 1.*

14.3.2 Scoring the Constructed-Response Items

As indicated in Table 14-5, about one-third of the geography items were constructed-response. Short constructed-response items were given a right/wrong scoring. The categories of responses for the items and the number of responses that were rescored for each item are indicated in Appendix H. The

BEST COPY AVAILABLE

percent agreement for the raters and Cohen's (1968) Kappa, a reliability estimate appropriate for items that are dichotomized, are also given in the tables. A 20 percent sample was used in calculating the rater reliability.

In general, the rater reliability of the scoring for dichotomized responses was quite high. Cohen's Kappa reliabilities ranged over items from .75 to .98 for grade 4, from .72 to .99 for grade 8, and from .69 to .98 for grade 12.

Chapter 7 discusses the definition of the item ratings and describes the process by which teams of raters scored the constructed-response items. This discussion includes the rating definitions for regular, short and extended constructed-response items as well as the range of interrater reliabilities that were obtained. Extended constructed-response items were scored on a scale from 1 to 5 to reflect degrees of knowledge. Rating information on extended constructed-response items can be found in Appendix H, which lists the sample sizes, percent agreement, and Cohen's Kappa reliability index.

14.3.3 Differential Item Functioning

A differential item functioning (DIF) analysis of the geography items was done to identify potentially biased items. Sample sizes were large enough to compare male and female students, White and Black students, and White and Hispanic students. The purpose of the analysis was to identify items that should be examined more closely by a committee of trained test developers and subject-matter specialists for possible bias and consequent exclusion from the assessment. The presence of DIF in an item means that the item is differentially harder for one group of students than another, while controlling for the ability level of the students.

For dichotomous items, the Mantel-Haenszel procedure as adapted by Holland and Thayer (1988) was used as a test of DIF (this is described in Chapter 9). The Mantel procedure (Mantel, 1963) was used for detection of DIF in polytomous items and also as described by Zwick, Donoghue, and Grima (1993). This procedure assumes ordered categories.

The grade-only portion of the geography sample was used for DIF analyses with weights rescaled separately for each comparison, as described in Chapter 9. DIF analyses were conducted separately by grade. For dichotomous items, the DIF index generated by the Mantel-Haenszel procedure was used to place items into one of three categories: "A," "B," or "C." "A" items exhibit no DIF, "B" items exhibit a weak indication of DIF, and "C" items exhibit a strong indication of DIF. "C" items were examined by a DIF committee for presence of bias.

Table 14-8 shows the count of dichotomous items with strong evidence of DIF. Focal groups are female, Black, and Hispanic groups. Positive values indicate items that were differentially easier for the focal group. Specifically, Table 14-8 indicates that there are only three dichotomous items that show strong evidence of differential item functioning and these all occur at grade 12. Of these, a) one item was more difficult for female students than for comparable male students, b) one item was easier for Black students than for comparable White students, and c) one item was easier for Hispanic students than for comparable White students. Table 14-9 shows items that had a standardized mean difference with an absolute value greater than a quarter of an item's standard deviation. Positive values were differentially easier for the focal group.

To aid in interpreting the Mantel statistic for polytomous items, the standardized mean difference between focal and reference groups was produced. This statistic was rescaled by dividing the standardized mean differences by the standard deviation of the respective item. The description of this procedure can be found in Chapter 9. For polytomous items, a standardized mean difference ratio of .25 or greater was considered a strong indication of DIF. It can be shown that standardized mean difference ratios of .25 are at least as extreme as Mantel-Haenszel statistics corresponding to "C" items (see Chapter 9 for details). In Table 14-9, items with a standardized mean difference ratio of at least .25 are listed under "CC," and all other DIF items are listed under "AA."

Table 14-8
DIF Category by Grade for Dichotomous Items

Grade	DIF Category*	Analysis		
		Male/Female	**White/Black**	**White/Hispanic**
4	C-	0	0	0
	B-	1	2	2
	A-	36	30	26
	A+	19	21	30
	B+	3	6	1
	C+	0	0	0
8	C-	0	0	0
	B-	4	4	9
	A-	47	34	35
	A+	30	38	32
	B+	3	8	8
	C+	0	0	0
12	C-	1	0	0
	B-	7	3	3
	A-	39	32	35
	A+	33	39	36
	B+	5	10	10
	C+	0	1	1

**Note: Positive values of the index indicate items that are differentially easier for the "focal" group (female, Black, or Hispanic students) than for the "reference" groups (male or White students). "A" means no indication of DIF, "B" means a weak indication of DIF, and "C" means a strong indication of DIF.*

BEST COPY AVAILABLE

Table 14-9

DIF Category by Grade for Polytomous Items

Grade	DIF Category	Analysis		
		Male/Female	White/Black	White/Hispanic
4	CC-	0	0	0
	AA-	3	20	17
	AA+	28	11	13
	CC+	0	0	1
8	CC-	1	1	0
	AA-	11	22	20
	AA+	26	18	20
	CC+	3	0	1
12	CC-	0	2	0
	AA-	13	27	15
	AA+	24	9	22
	CC+	1	0	1

***Note:** Positive values of the index indicate items that are differentially easier for the "focal" group (female, Black, or Hispanic students) than for the "reference" groups (male or White students). "AA" means little or no indication of DIF and "CC" means a strong indication of DIF.*

Table 14-9 describes DIF for polytomous items. Only CC+ (DIF in favor of focal group) and CC- (DIF against focal group) items were considered to show strong evidence of DIF. Grade 4 had one item with a strong indication of DIF, while at grade 8 there were six such items, and at grade 12 there were four. In all, 90 items were analyzed for DIF at grade 4, while 125 items were analyzed at grade 8, and 123 items at grade 12.

Following standard practice at ETS, all items identified as showing DIF were reviewed by a committee of trained test developers and subject-matter specialists. As described in Chapter 9, such committees are charged with making judgments about whether the differential difficulty of an item is *unfairly* related to group membership; that is, whether the item is biased. The committee assembled to review NAEP items included both ETS staff and outside members with expertise in the field. It was the committee's judgment that none of the DIF items for the geography assessment data were functioning differentially due to factors irrelevant to test objectives; that is, none were biased against particular groups of students. Hence, none of the items were removed from scaling. It is important to note that if all differentially functioning items were automatically deleted without an independent assessment of the presence of bias, the result could obscure legitimate proficiency differences between subgroups and possibly alter valid policy interpretations of NAEP results.

BEST COPY AVAILABLE

14.4 IRT SCALING

14.4.1 Item Parameter Estimation

The NAEP BILOG/PARSCALE computer program was used to estimate the item parameters for the assessment. For multiple-choice and dichotomized short constructed-response items, a three-parameter IRT model was used. Regular and extended constructed-response items were polytomously scored and were analyzed with a generalized partial credit model (Muraki, 1992). Although geography items were not written according to a cross-grade framework, and the items from all grades were not scaled together, the three grade-only scales were concatenated to a metric with the appearance of a cross-grade scale in accordance with the wishes of the National Center for Education Statistics (NCES). This was accomplished through the use of the TBLT program (Muraki & Grima, 1993), which provides the constants that transform student proficiency scores to a common cross-grade metric. (See Chapter 9 for further description of transformation to the reporting metric). Within each grade, items were scaled using the full age/grade sample of students available from the 1994 assessment. Item parameters were estimated separately for each of the three content area scales. The final BILOG/PARSCALE estimation of item parameters did not assume that the population ability distribution had a specific shape or mean. (See Chapter 9 for further descriptions of the scaling process). The calibration used student weights that were rescaled so that their sum equaled the unweighted sample size.

A complexity introduced by the 50-minute blocks in geography is that those blocks of items must be linked in some way to the shorter blocks. This is complicated by the fact that no students received the shorter blocks in addition to the 50-minute blocks. Because the samples of students receiving each booklet are representative of the population as a whole, it was assumed that the distribution of student proficiency was the same for the students receiving the 50-minute blocks as for the students receiving the booklets containing the shorter blocks.

14.4.2 Evaluation of Model Fit

During the course of estimating an IRT model, individual items were evaluated to determine how well the item response model fit the data. This was done by visual inspection of plots comparing empirically based and theoretical item response functions (see Chapter 9 for a fuller explanation of these plots). For most items, the model fit well. In a few cases, poor fit with the data led to special treatment or deletion of the item.

Several items received special treatment. Of the 338 items across grades (counting overlaps separately for each grade), 12 (four percent) received special treatment. These items are listed in Table 14-10, along with the reason for special treatment. Six dichotomous items with a nonmonotonic or nonlogistic item response function were dropped from the assessment. Six were polytomous items that had poor model fit and were dichotomized (i.e., categories were collapsed), resulting in two-category items.

In making decisions about excluding items from the final scales or collapsing categories of polytomous items, a balance was sought between being too stringent, hence deleting or redefining too many items and possibly damaging the content representativeness of the pool of scaled items, and being too lenient, hence including items with model fit poor enough to endanger the types of model-based inferences made from NAEP results. Items that clearly did not fit the model were excluded or recoded in the final scales; however, a certain degree of misfit was tolerated for a number of items included in the final scales.

In order to convey a sense of what it means for items to fit the data, example item plots are presented that show theoretical item characteristic curves and empirical item responses. The X-axis represents geography proficiency while the Y-axis is the probability of getting a correct response on the item. For dichotomous items, the solid lines show how probable it is that a student, at a certain level of proficiency, will get an item correct. For polytomous items, the solid line corresponding to a certain category indicates how probable it is that a student at a certain level of proficiency will score in that category. The line composed of diamonds represents the empirical item function (i.e., it approximates actual behavior of the students). The size of the diamonds represents the proportion of the sample at a certain ability (θ) level. A fuller explanation of item plots can be found in Chapter 9.

An item fits the data well if the empirical trace (indicating observed behavior of students) and the theoretical curve (indicating theoretically predicted student behavior) are close together. Figure 14-1 shows a dichotomous item that fits the data well, as indicated by the empirical curve (diamonds) and the theoretical curve (solid line) being close together. Figure 14-2 shows a poor fitting item—the empirical and theoretical lines go in different directions. This item was deleted from the assessment. Figure 14-3 shows a poor fitting polytomous item; the empirical traces diverge from the theoretical lines for every category. In Figure 14-4 the same item has been collapsed so that it is dichotomous (the last two categories of three categories were collapsed). Figure 14-5 illustrates a four category polytomous item which fits the data reasonably well.

Note that the plots shown in Figures 14-1 through 14-5 were obtained from preliminary item parameter calibrations. They are presented to reflect the information used to make the decisions discussed in the text. Plots produced from the final item parameters (listed in Appendix D) were very similar to those presented and supported the decisions made.

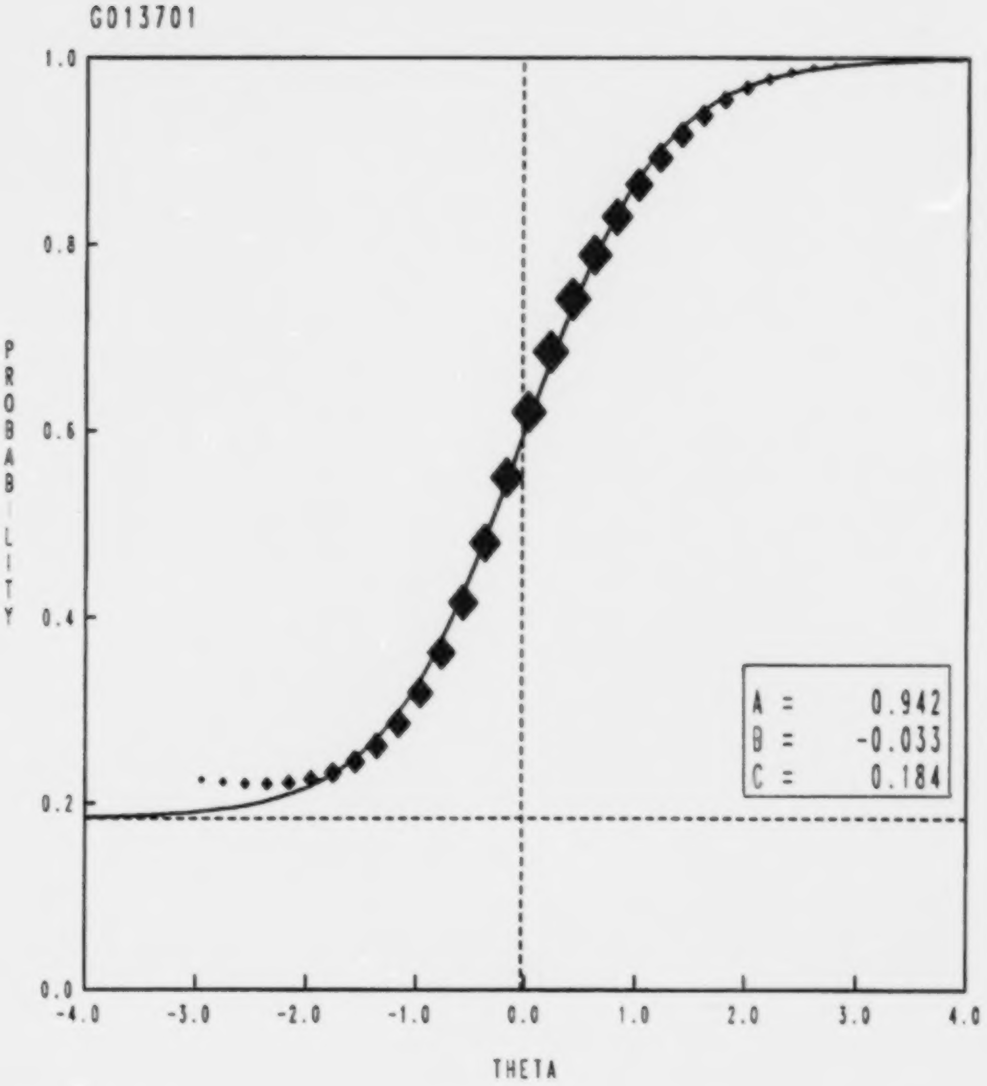
Table 14-10

Items from the 1994 Geography Main Analysis Receiving Special Treatment

Grade	Item	Scale	Block	Type	Problem	Disposition
4	G012201	Space and Place	G8	Polytomous	Category 2 almost empty.	Dichotomized, Categories 2 and 3 (of 3) collapsed.
	G009701	Environment and Society	G5	Dichotomous	Nonmonotonic item response function.	Dropped.
	G012902	Spatial Dynamics	G8	Polytomous	Category 2 almost empty.	Dichotomized, Categories 2 and 3 (of 3) collapsed.
8	G016201	Space and Place	G5	Polytomous	Bad fit.	Dichotomized, Categories 1 and 2 (of 3) collapsed.
	G017201	Space and Place	G5	Dichotomous	Nonmonotonic item response function.	Dropped.
	G012503	Spatial Dynamics	G8	Polytomous	Bad fit.	Dichotomized, Categories 2 and 3 (of 3) collapsed.
	G012902	Spatial Dynamics	G8	Polytomous	Category 2 almost empty.	Dichotomized, Categories 2 and 3 (of 3) collapsed.
12	G017201	Space and Place	G5	Dichotomous	Nonmonotonic item response function.	Dropped.
	G026502	Space and Place	G9	Polytomous	Bad fit	Dichotomized, Categories 1 and 2 (of 3) collapsed.
	G027101	Space and Place	G9	Dichotomous	Non-logistic item response function.	Dropped.
	G021201	Environment and Society	G3	Dichotomous	Nonmonotonic item response function.	Dropped.
	G022701	Spatial Dynamics	G4	Dichotomous	Nonmonotonic item response function.	Dropped.

Figure 14-1

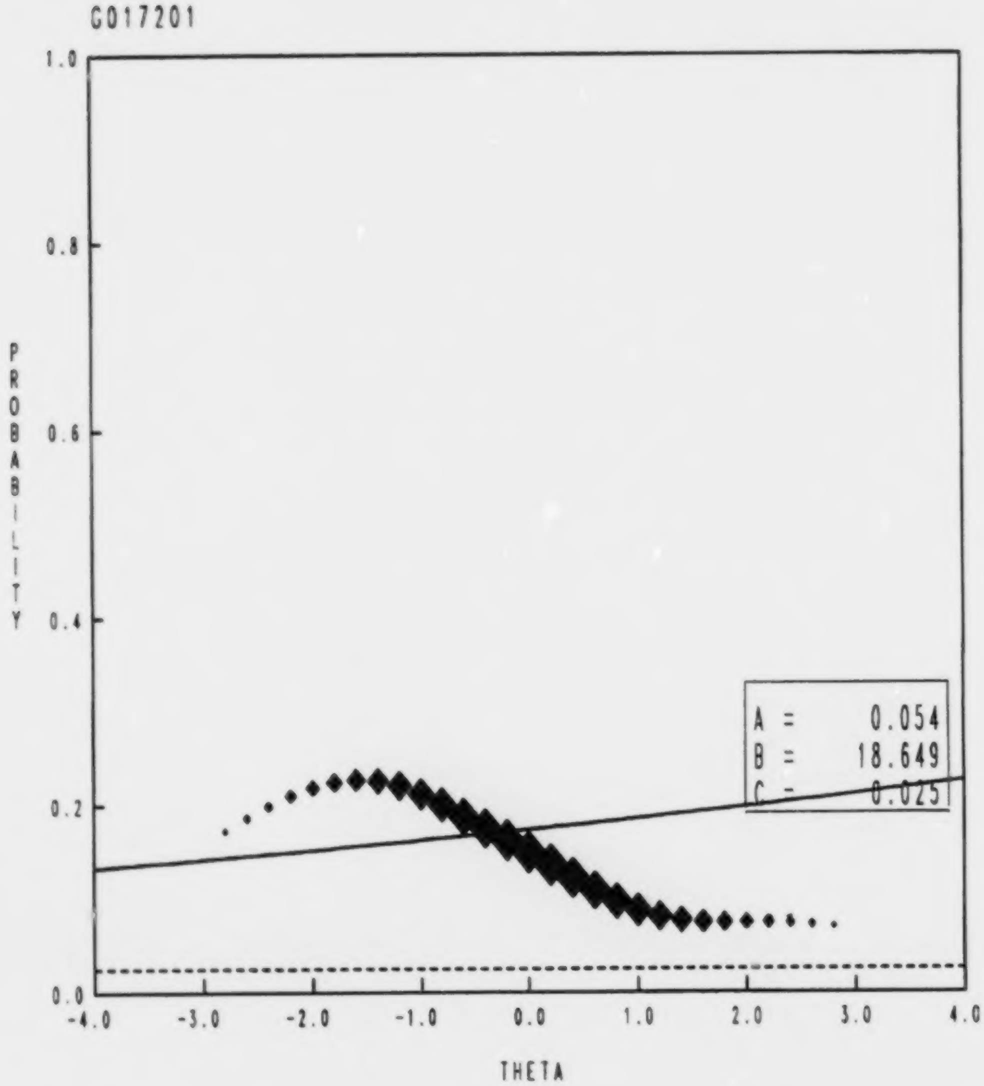
Example Dichotomous Item (G013701)
Exhibiting Good Model Fit



***Note**: The plot compares empirical and model-based estimates of the item response function (IRF). The smooth curve represents the model-based estimate at each provisional proficiency level. The diamonds represent the empirical item function (i.e., it approximates actual behavior of the students), and the size of the diamonds represents the proportion of the sample at a certain ability (θ) level.*

Figure 14-2

Example Dichotomous Item (G017201)
Exhibiting Poor Model Fit

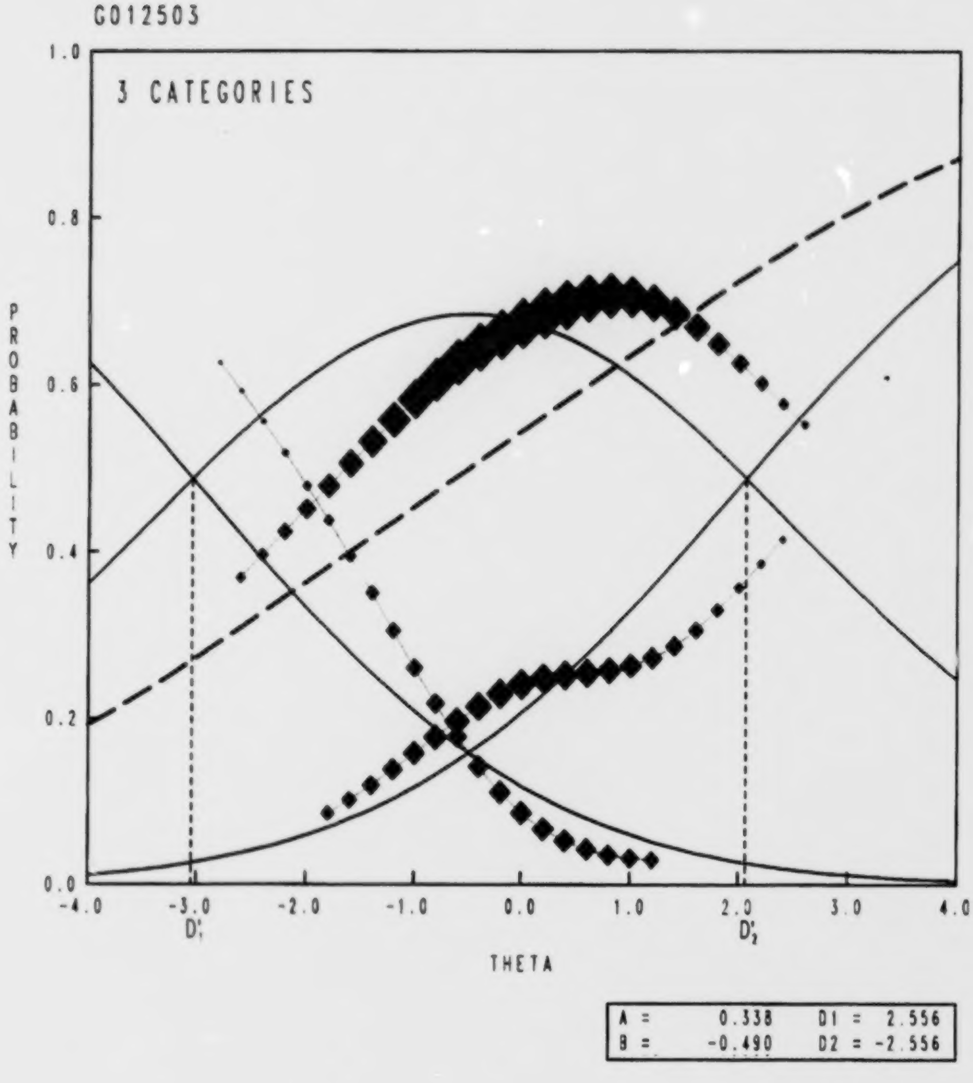


Note*: *The plot compares empirical and model-based estimates of the item response function (IRF). The smooth curve represents the model-based estimate at each provisional proficiency level. The diamonds represent the empirical item function (i.e., it approximates actual behavior of the students), and the size of the diamonds represents the proportion of the sample at a certain ability (θ) level.*

BEST COPY AVAILABLE

Figure 14-3

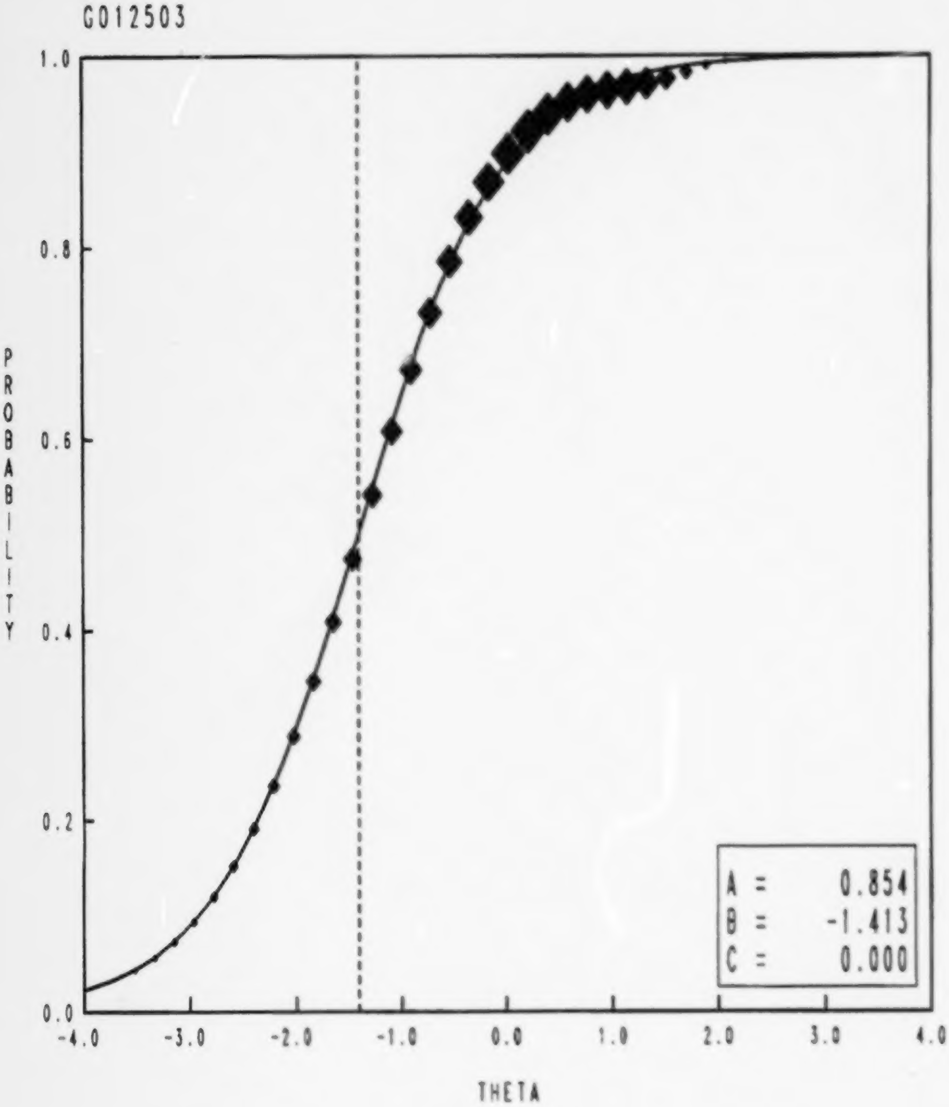
Example Polytomous Item (G012503)
Exhibiting Poor Model Fit



***Note:** The plot compares empirical and model-based estimates of the item response function (IRF). The smooth curve represents the model-based estimate at each provisional proficiency level. The diamonds represent the empirical item function (i.e., it approximates actual behavior of the students), and the size of the diamonds represents the proportion of the sample at a certain ability (θ) level.*

Figure 14-4

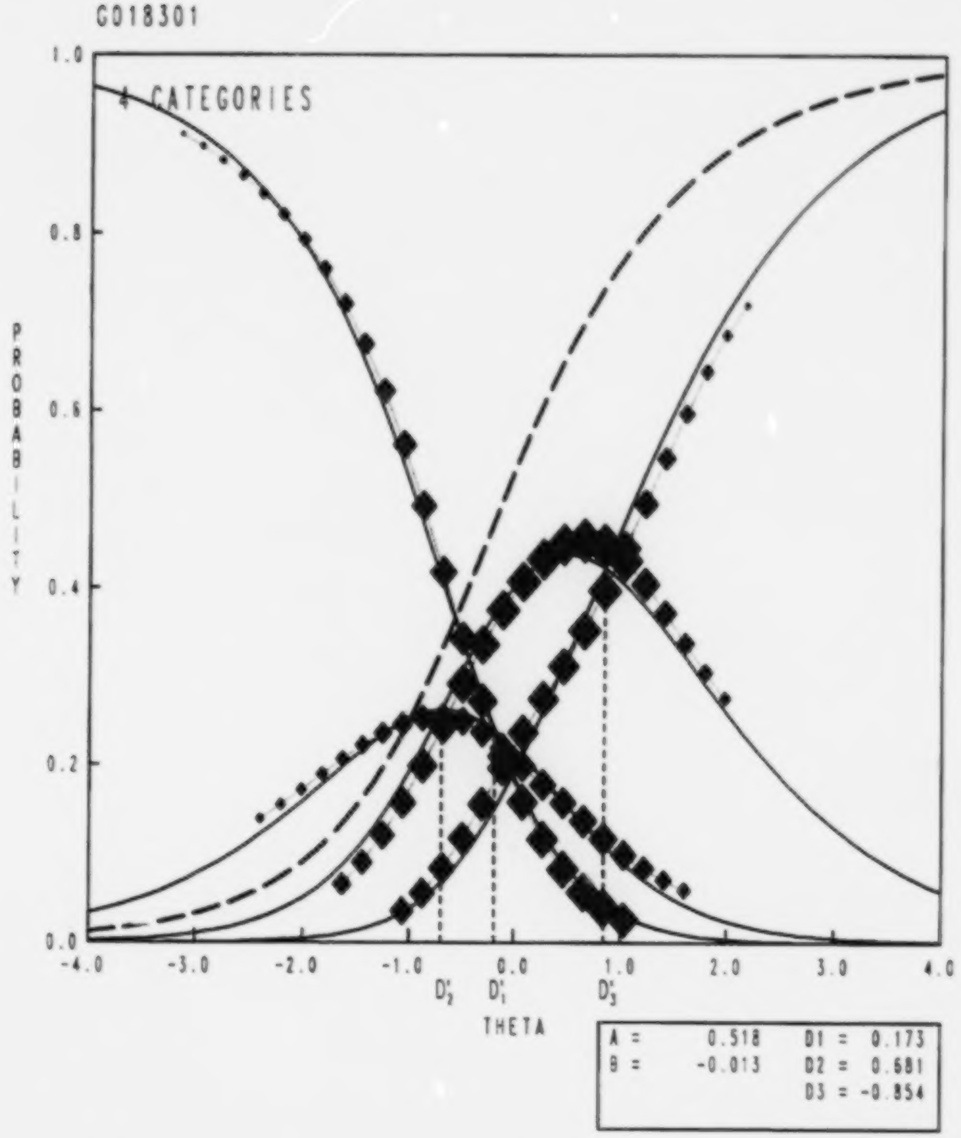
Example Dichotomous Item (G012503) Exhibiting Good Model Fit After Polytomous Categories (Seen in Figure 14-3) Were Collapsed



****Note:*** *The plot compares empirical and model-based estimates of the item response function (IRF). The smooth curve represents the model-based estimate at each provisional proficiency level. The diamonds represent the empirical item function (i.e., it approximates actual behavior of the students), and the size of the diamonds represents the proportion of the sample at a certain ability (θ) level.*

Figure 14-5

Example Polytomous Item (G018301)
Exhibiting Good Model Fit



***Note**: *The plot compares empirical and model-based estimates of the item response function (IRF). The smooth curve represents the model-based estimate at each provisional proficiency level. The diamonds represent the empirical item function (i.e., it approximates actual behavior of the students), and the size of the diamonds represents the proportion of the sample at a certain ability (θ) level.*

14.5 GENERATION OF PLAUSIBLE VALUES

14.5.1 Principal Components (NSWEEP Program)

Multivariate plausible values were generated for the entire age/grade sample using the multivariate conditioning program CGROUP as revised by Thomas (1994). This procedure employed student weights. Prior to 1990, selected background variables were used for conditioning. However, from 1990 to the present, principal components of the background variables have been used as conditioning variables. Principal components were employed to remedy problems of extreme collinearity among some of the original conditioning variables. The principal components used accounted for 90 percent of the variance of the original conditioning variables.

Results from research on the 1990 Trial State Assessment in mathematics suggests that using a large subset of principal components will yield estimates that differ only slightly from those obtained using the full set (Mazzeo, Johnson, Bowker, & Fong, 1994). Table 14-11 contains a list of the number of principal components included in conditioning, as well as the proportion of variance accounted for by the conditioning model for each age/grade.

Table 14-11
Proportion of Proficiency Variance Accounted for by the Conditioning Model for the Geography Main Assessment

Grade	Number of Conditioning Contrasts	Number of Principal Components	Proportion of Proficiency Variance Accounted for		
			Space and Place	Environment and Society	Spatial Dynamics and Connections
4	571	232	.70	.67	.65
8	553	219	.58	.59	.60
12	403	174	.65	.67	.67

14.5.2 Conditioning (CGROUP Program)

The codings of the original geography-specific conditioning variables, before principal components were calculated, are presented in Appendix C. (For age 17/grade 12, the "modal age, > modal grade" category was deleted from the age-by-grade variable, because students above grade 12 were not sampled.) NAEP-CGROUP (described in Chapter 11) creates posterior distributions of proficiencies by combining information from item responses of individuals and information from linear regression of proficiency on conditioning variables. For each individual, five plausible values are randomly drawn from their posterior proficiency distribution.

The values of the conditioning effects are expressed in the metrics of the original calibration scale. Definitions of derived conditioning variables are given in Appendix B.

14.5.3 Analysis of Dimensionality

As mentioned earlier, the main assessment is multivariate with three content area scales. Tables 14-12 and 14-13 give conditional and marginal correlations for the three scales for the three grades. The conditional correlations can be thought of as correlations from information pooled within the demographic subgroups corresponding to grouping variables used to condition the data with CGROUP. The conditional correlations correspond to the error correlations of a CGROUP analysis. The conditional correlations are quite high, averaging .90 for grade 4, .94 for grade 8, and .89 for grade 12. The marginal correlations (computed on a grade only sample) are the average of the scale correlations over five plausible values generated by CGROUP. Since these correlations are not pooled within background groupings, marginal correlations tend to be larger than conditional correlations, averaging .94 for grade 4, .95 for grade 8, and .93 for grade 12. Although it is of substantive interest to analyze the scales separately, the correlations indicate that they are highly redundant with each other.

Table 14-12

Conditional Correlations of Geography from Conditioning (CGROUP)*

Age/ Grade	Scale	Space and Place	Environment and Society	Spatial Dynamics and Connections
9/4	Space and Place	1.00	–	–
	Environment and Society	.86	1.00	–
	Spatial Dynamics and Connections	.91	.94	1.00
13/8	Space and Place	1.00	–	
	Environment and Society	.92	1.00	–
	Spatial Dynamics and Connections	.94	.95	1.00
17/12	Space and Place	1.00	–	–
	Environment and Society	.85	1.00	–
	Spatial Dynamics and Connections	.88	.94	1.00

****Note***: *These correlations are based on the age and grade sample.*

Table 14-13

*Marginal Correlations of Geography Scales**

Grade	Scale	Space and Place	Environment and Society	Spatial Dynamics and Connections
4	Space and Place	1.00	–	–
	Environment and Society	.91	1.00	–
	Spatial Dynamics and Connections	.94	.94	1.00
8	Space and Place	1.00	–	
	Environment and Society	.94	1.00	–
	Spatial Dynamics and Connections	.94	.95	1.00
12	Space and Place	1.00	–	–
	Environment and Society	.92	1.00	–
	Spatial Dynamics and Connections	.93	.95	1.00

****Note:*** *These correlations are based on the "grade only" portion of each sample. Tabled values were obtained by computing a separate Pearson correlation coefficient for each plausible value, computing Fisher's z-transformation for each value, computing the average of the transformed values, and computing the inverse transformation of the average.*

14.6 THE FINAL GEOGRAPHY SCALES

14.6.1 Linking of the Within-Grade Scales Using the TBLT Program

Like all IRT scales, the within-grade geography scales have a linear indeterminacy that may be resolved by an arbitrary choice of the origin and unit-size standard deviation in each given scale. By convention, for each scale, the geography proficiencies were to be put on a cross-grade scale with a mean of 250 and a standard deviation of 50. The TBLT program (Muraki & Grima, 1993) was used to find a transformation to put the grade 4 thetas (or ability scores) on the grade 8 scale and similarly to put the grade 12 thetas on the grade 8 scale. In a second step, the thetas on the cross-grade scale were transformed to the arbitrary 250, 50 scale. In application, these two steps were combined. Table 14-14 gives the constants to linearly transform the thetas from each within-grade scale to a cross-grade scale. The transformation had the form

$$\theta_{\text{proficiency}} = A \cdot \theta_{\text{calibrated}} + B,$$

with $\theta_{\text{calibrated}}$ being the ability estimate on a within-grade scale and $\theta_{\text{proficiency}}$ being the ability estimate on the cross-grade that has a mean of 250 and a standard deviation of 50.

Table 14-14

Coefficients of Linear Transformations of the Geography Scales from the Calibrating Scale Units to the Units of Reporting Proficiency

Age/Grade	Scale	A	B
9/4	Space and Place	45.50	200.09
	Environment and Society	49.68	195.64
	Spatial Dynamics and Connections	43.87	196.35
13/8	Space and Place	38.64	254.99
	Environment and Society	38.64	257.73
	Spatial Dynamics and Connections	38.20	252.30
17/12	Space and Place	35.39	279.16
	Environment and Society	28.14	278.45
	Spatial Dynamics and Connections	31.30	283.51

14.6.2 The Composite Geography Scales

While multiple scales provide useful and very revealing information about the relative relationships among subpopulations, a single index to summarize overall performance is useful and communicative. For that reason, a geography composite was defined as a weighted average of the results across content area scales. The emphasis associated with each content area scale remains the same across grades. The assigned weights reflect the relative importance of content area scales for a particular age/grade as specified in the objectives for the geography assessment (NAGB, 1993). The weights for the composite for each age/grade sample is given in Table 14-15.

Table 14-15

Weights Used for Each Scale to Form the Geography Composite

Scale	Age 9/Grade 4	Age 13/Grade 8	Age 17/Grade 12
Space and Place	40	40	40
Environment and Society	30	30	30
Spatial Dynamics and Connections	30	30	30

The mean and standard deviations of the composite scales for all three grades is given in Table 14-16.

Table 14-16

Means and Standard Deviations on the Geography Composite Scale

Grade	Mean*	S.D.
4	205.7	43.9
8	259.7	34.9
12	284.6	29.7

**Note: Means were an average of all five plausible values.*

14.7 PARTITIONING OF THE ESTIMATION ERROR VARIANCE

For each scale within each grade, the error variance of the transformed proficiency mean was partitioned according to the procedure described in Chapter 9. The variance is partitioned into two parts; the proportion of error variance due to sampling students (sampling variance) and the proportion of error variance due to the fact that proficiency, θ, is a latent variable that is estimated rather than observed. Table 14-17 contains estimates of the total error variance, the proportion of error variance due to sampling students and the proportion of error variance due to the latent nature of θ (for stability of the estimates, these are based on 100 plausible values). More detailed information by gender and race/ethnicity is presented in Appendix E.

Table 14-17

Estimation Error Variance and Related Coefficients for the Geography Main Assessment

Grade	Scale	Total Estimation Error Variance	Proportion of Variance Due to ...	
			Student Sampling	Latency of θ
4	Space and Place	1.419	0.92	0.08
	Environment and Society	1.492	0.87	0.13
	Spatial Dynamics and Connections	1.444	0.83	0.17
	Composite	1.265	0.95	0.05
8	Space and Place	0.887	0.90	0.10
	Environment and Society	0.776	0.90	0.10
	Spatial Dynamics and Connections	0.983	0.92	0.08
	Composite	0.722	0.97	0.03
12	Space and Place	0.871	0.92	0.08
	Environment and Society	0.794	0.88	0.12
	Spatial Dynamics and Connections	0.890	0.90	0.10
	Composite	0.731	0.96	0.04

14.8 GEOGRAPHY TEACHER QUESTIONNAIRES

Teachers of fourth- and eighth-grade students were surveyed about their educational background and teaching practices. The students in a particular classroom had their records matched with their teacher's survey information. This was done for the "grade-only" sample. Variables derived from the questionnaire were used in the conditioning models, along with a variable that indicated whether a student record had been matched with a teacher record, which controls estimates of subgroup means for differences that exist between the matching and non-matching students. Of the 5,507 fourth-grade students in the sample, 89 percent were matched with both parts of the teacher questionnaire and seven percent were matched with only the first, teacher background, part of the questionnaire. For the eighth-grade students sample, 42 percent were matched with both parts of the teacher questionnaire and 37 percent were matched with only the first part of the questionnaire. The lower match rate for both parts of the questionnaire for eighth-grade students was due in part to the fact that in grade 8 students were matched to the particular class that the teacher taught. Class membership information was often missing or ambiguous. For grade 4 students only had to be matched to the teacher, resulting in higher match rates. Thus, 95 percent of the fourth graders and 80 percent of the eighth graders were matched with at least the background information about their geography teachers.

Chapter 15

DATA ANALYSIS FOR THE LONG-TERM TREND READING ASSESSMENT[1]

Hua-hua Chang, John R. Donoghue, Lois H. Worthington, Minhwei Wang, and David S. Freund
Educational Testing Service

15.1 INTRODUCTION

This chapter describes the analyses performed on the responses to the cognitive and background items in the 1994 assessment of reading. These analyses led to the results presented in the *NAEP 1994 Trends in Academic Progress: Achievement of U.S. Students in Science, 1969 to 1994; Mathematics, 1973 to 1994; Reading, 1971 to 1994; and Writing, 1984 to 1994* (Campbell, Reese, O'Sullivan, Dossey, & Donahue, 1996). The emphasis of this chapter is on the methods and results of procedures used to develop the IRT-based scale scores that formed the basis of this report. The theoretical underpinnings of the IRT and plausible values methodology described in this chapter are given in Chapter 11, and several of the statistics are described in Chapter 9.

The objectives of the reading analyses were to prepare scale values and perform all analyses necessary to produce a long-term trend report in reading. The reading long-term trend results include the years 1971, 1975, 1980, 1984, 1988, 1990, 1992, and 1994. The major analysis components are discussed in turn. Some aspects of the analysis, such as procedures for item analysis, scoring of constructed-response items, and methods of scaling, are described in previous chapters and are therefore not detailed here.

The student samples that were administered reading items in the 1994 long-term trend reading assessment are shown in Table 15-1. (See Chapters 1 and 3 for descriptions of the target populations and the sample design used for the assessment.) Data from the long-term trend samples that contributed to the trends in reading achievement were scaled separately from the 1994 main focused-BIB samples. Accordingly, the long-term trend and main analyses are presented in separate chapters. This chapter pertains to the scaling of the long-term trend data; information about the scaling of the data from the main focused-BIB samples is presented in Chapter 12.

---

[1] Hua-hua Chang was the primary person responsible for the planning, specification, and coordination of the reading long-term trend analyses. John R. Donoghue advised and reviewed the scaling of the reading long-term trend data. Computer activities for all long-term trend reading scaling and data analyses were directed by Minhwei Wang and performed by Lois H. Worthington and David S. Freund.

Table 15-1

NAEP 1994 Long-Term Trend Reading Student Samples

Sample	Booklets	Mode	Cohort Assessed	Time of Testing	Age Definition	Modal Grade	Number Assessed
9 [RW-LTTrend]	51-56	Print	Age 9/ Grade 4	Winter	CY	4	5,335
13 [RW-LTTrend]	51-56	Print	Age 13/ Grade 8	Fall	CY	8	5,547
17 [RW-LTTrend]	51-56	Print	Age 17/ Grade 11	Spring	Not CY	11	4,840

LEGEND

RW	*Reading and writing*
LTTrend	*Long-term trend assessment*
Print	*Printed administration*
CY	*Calendar year: birthdates in 1984, 1980, and 1976 for ages 9, 13, and 17*
Not CY	*Age 17 only: birthdates between October 1, 1976, and September 30, 1977*

The trend results reported in *Trends in Academic Progress* are based on print administrations and occur at all of the age levels. The samples involved in the analysis are shown in Table 15-1 as 9[RW-LTTrend], 13[RW-LTTrend], and 17[RW-LTTrend]. The long-term trend booklets for these samples contained blocks of reading and writing items administered in print form. All students received a block of common background questions, distinct for each age, and several subject-area background questions, which were presented in the cognitive blocks. The booklets are identical to those used for long-term trend assessments in 1984, 1988, 1990, and 1992. The booklets and the blocks within those booklets are listed in Chapter 4. Additional information about all of the items in these blocks also appears in that chapter. This chapter includes specific information about the long-term trend items that were scaled. Both age- and grade-selected students contributed to the long-term trend scaling. However, only students in the "age-only" portion of the reading long-term trend samples contributed to the results presented in *Trends in Academic Progress*.

Table 15-2 clarifies the relationships between the 1994 long-term trend samples and samples from previous years. For ages 9, 13, and 17, the [RW-LTTrend] samples allow direct comparisons with 1992, 1990, 1988, and 1984 samples. The long-term trend scale, established in 1984, was linked to the 1971, 1975, and 1980 assessments using a complex equating strategy described in *Implementing the New Design: The NAEP 1983-84 Technical Report* (Beaton, 1987). At each age, several booklets were retained intact from the 1984 assessment, forming the basis of the reading long-term trend assessment in 1988, 1990, 1992, and 1994. Information about the 1988 assessment is available in *Focusing the New Design: The NAEP 1988 Technical Report* (Johnson & Zwick, 1990); information about the 1990 assessment is given in *The NAEP 1990 Technical Report* (Johnson & Allen, 1994); and information about the 1992 assessment is given in *The NAEP 1992 Technical Report* (Johnson & Carlson, 1994).

Table 15-2

NAEP Reading Samples Contributing to 1994 Long-Term Trend Results, 1971-1994

Cohort	Year	Sample	Subjects	Time of Testing	Mode	Age Definition	Modal Grade
Age 9	1971	Main	RL	Winter	Tape	CY	4
	1975	Main	RA	Winter	Tape	CY	4
	1980	Main	RA	Winter	Tape	CY	4
	1984	Main	RW	Winter, spring	Print	CY	4
	1984	T-84	RW	Winter	Tape	CY	4
	1988	LTTrend	RW	Winter	Print	CY	4
	1990	LTTrend	RW	Winter	Print	CY	4
	1992	LTTrend	RW	Winter	Print	CY	4
	1994	LTTrend	RW	Winter	Print	CY	4
Age 13	1971	Main	RL	Fall	Tape	CY	8
	1975	Main	RA	Fall	Tape	CY	8
	1980	Main	RA	Fall	Tape	CY	8
	1984	Main	RW	Winter, spring	Print	CY	8
	1984	T-84	RW	Fall	Tape	CY	8
	1988	LTTrend	RW	Fall	Print	CY	8
	1990	LTTrend	RW	Fall	Print	CY	8
	1992	LTTrend	RW	Fall	Print	CY	8
	1994	LTTrend	RW	Fall	Print	CY	8
Age 17	1971	Main	RL	Spring	Tape	Not CY	11
	1975	Main	RABS	Spring	Tape	Not CY	11
	1980	Main	RA	Spring	Tape	Not CY	11
	1984	Main	RW	Winter, spring	Print	Not CY	11
	1984	T-84	RW	Spring	Tape	Not CY	11
	1988	LTTrend	RW	Spring	Print	Not CY	11
	1990	LTTrend	RW	Spring	Print	Not CY	11
	1992	LTTrend	RW	Spring	Print	Not CY	11
	1994	LTTrend	RW	Spring	Print	Not CY	11

******Note*: *Withan an age class, these samples received common booklets.*

LEGEND

RL	*Reading and literature*
RA	*Reading and art*
RABS	*Reading, art, index of basic skills*
RW	*Reading and writing*
Main	*Main assessment*
T-84	*Special sample in the 1984 assessment that was used to establish links to previous assessments (1971-1980) for the purposes of long-term trend*
LTTrend	*Long-term trend (these samples received common booklets within an age group)*
Print	*Print administration*
Tape	*Audiotape administration*
CY	*Calendar year: birthdates (1994 sample) in 1984 and 1980 for ages 9 and 13*
Not CY	*Age 17 only (1994 sample): birthdates between October 1, 1976, and September 30, 1977*

BEST COPY AVAILABLE

The 1994 long-term trend included, at each age level, six of the assessment booklets administered in 1984. These booklets (51-56) contained both reading and writing blocks, as well as background items. Although these long-term trend booklets represented only about one-tenth of the reading booklets administered in the complex 1984 BIB design,[2] they contained 10 of the 12 reading blocks that were scaled at each age/grade level in 1984. The samples of students who received these long-term trend booklets are described in Table 15-1 and in Chapter 3. The purpose of the reading long-term trend analysis was to add to the reading trend results that extended from 1971 to 1992 for ages 9, 13, and 17. The numbers of scaled items for each age are presented in Table 15-3. Each age was scaled separately. The numbers of items scaled in 1994 that were common across assessment years are given in Table 15-4. As was the case for previous long-term trend analyses, the long-term trend scale is univariate. Dimensionality analyses conducted following the 1984 assessment showed that the reading items were well summarized by a unidimensional scale (Zwick, 1987).

Table 15-3
Numbers of Scaled Reading Long-Term Trend Items Common Across Ages

Age	Number of Items
9 only	60
13 only	21
17 only	23
9 and 13 only	13
9 and 17 only	2
13 and 17 only	42
9, 13, and 17	27
Total	**188**

---

[2] The long-term trend assessment included 1984 booklets 16, 17, 27, 34, 55, and 60 at age 9 and booklets 13, 16, 17, 21, 34, and 57 at ages 13 and 17 (see J. R. Johnson, 1987, pp. 120-121). The 1984 main assessment focused BIB design included 57 booklets that contained at least one scaled reading block at age 9 and 56 such booklets at ages 13 and 17.

Table 15-4

Numbers of Scaled Reading Long-Term Trend Items Common Across Assessments

Assessment Year	Number of Items		
	Age 9	Age 13	Age 17
1984, 1990, 1992, 1994	101	101	93
1984, 1992, 1994	102	103	94
1984, 1988, 1990, 1992, 1994	98	98	87
1980, 1984, 1988, 1990, 1992, 1994	67	71	52
1971, 1975, 1980, 1984, 1988, 1990, 1992, 1994	36	45	37

The steps in the reading long-term trend analysis are documented in the following sections. As is usual in NAEP analyses, the first step was to gather item and block information. The long-term trend items were then calibrated according to the IRT model. Plausible values were generated after conditioning on available background variables. Finally, the scale values were placed on the final reading long-term trend scale used in previous trend assessments.

15.2 ITEM ANALYSIS FOR THE READING LONG-TERM TREND ASSESSMENT

Conventional item analyses did not identify any difficulties with the long-term trend data. Table 15-5 contains the number of items, size of the sample administered the block, mean weighted proportion correct, mean weighted r-biserial, and mean weighted alpha as a measure of reliability for each block. Because the blocks were presented in self-paced, print-administered form, the weighted proportion of students attempting the last item is included in the table to give an indication of the speededness of each block. Common labeling of these blocks across ages does not denote common items. The correspondence between blocks, booklets, and samples is given for the reading long-term trend assessment in Chapter 4. Student weights were used for all statistics, except for the sample sizes. The average values reflect only the items in the block that were scaled. The average weighted proportion correct for age 13 and 17 tended to be slightly lower than it was for 1992 (see Table 12-5 in Johnson & Carlson, 1994). Overall, however, the 1994 item-level statistics were not very different from those for the 1984, 1988, 1990, and 1992 assessments.

Table 15-5

Descriptive Statistics for Item Blocks in the Reading Long-Term Trend Samples

AGE 9											
Statistics	Blocks										
	BH	BJ	BK	BL	BM	BN	BO	BP	BQ	BR	BV
Number of scaled items	10	8	11	7	11	12	11	–	11	12	9
Number of scaled constructed-response items	1	0	0	1	1	1	0	–	0	0	3
Unweighted sample size	627	662	662	649	614	614	653	–	649	1293	666
Average weighted proportion correct	.62	.54	.43	.50	.42	.56	.49	–	.54	.47	.63
Average weighted r-biserial	.77	.72	.65	.82	.67	.74	.60	–	.71	.66	.77
Weighted alpha reliability	.78	.68	.75	.75	.73	.83	.66	–	.80	.76	.78
Weighted proportion of students attempting last item	.91	.91	.79	.70	.67	.65	.88	–	.81	.83	.98

AGE 13											
Statistics	Blocks										
	BH	BJ	BK	BL	BM	BN	BO	BP	BQ	BR	BV
Number of scaled items	12	9	8	5	11	12	10	9	16	11	–
Number of scaled constructed-response items	1	0	0	0	1	1	1	1	0	0	–
Unweighted sample size	668	657	675	694	675	641	654	657	694	668	–
Average weighted proportion correct	.63	.62	.65	.72	.58	.67	.65	.73	.61	.69	–
Average weighted r-biserial	.69	.67	.73	.88	.67	.70	.62	.79	.58	.77	–
Weighted alpha reliability	.69	.62	.67	.60	.70	.79	.52	.70	.72	.79	–
Weighted proportion of students attempting last item	.96	.94	.98	.98	.93	.75	.84	.90	.77	.97	–

Table 15-5 (continued)

Descriptive Statistics for Item Blocks Reading Long-Term Trend Samples

AGE 17											
	Blocks										
Statistics	BH	BJ	BK	BL	BM	BN	BO	BP	BQ	BR	BV
Number of scaled items	12	54	8	6	11	12	13	10	10	7	–
Number of scaled constructed-response items	1	1	0	1	1	1	1	1	0	0	–
Unweighted sample size	658	652	644	608	644	654	595	652	608	658	–
Average weighted proportion correct	.71	.81	.78	.73	.68	.82	.67	.75	.57	.68	–
Average weighted r-biserial	.76	.98	.83	.83	.79	.80	.59	.80	.62	.84	–
Weighted alpha reliability	.72	.63	.69	.46	.73	.80	.71	.79	.64	.77	–
Weighted proportion of students attempting last item	.95	.96	.99	.96	.94	.88	.64	.84	.93	.97	–

15.3 TREATMENT OF CONSTRUCTED-RESPONSE ITEMS

Data for constructed-response items in the long-term trend analysis were used for the 1984, 1990, 1992, and 1994 assessments only. Constructed-response items were not included in the original scoring of the 1988 reading assessment because a previous study (Zwick, 1988) had shown that scoring inconsistencies (drops in interrater reliability and/or scorer drift—that is, scorers showing evidence of rating items more strictly or more leniently than did the original 1984 scorers) had affected these items. A similar review was performed on the 1994 constructed-response items. In general, the 1994 scoring did not suffer from the same inconsistencies as the 1988 scoring. Therefore, most of the 1994 constructed-response items were used in the long-term trend analysis.

At each age, several constructed-response items were found to exhibit drops in interrater reliability and/or scorer drift. Items exhibiting marked item drift or drops in interrater reliability were excluded from calibration. These items are listed in Table 15-6. The remaining constructed-response items were dichotomized according to criteria developed by subject-area experts. The dichotomized versions of the constructed-response items were included in the calibration.

Table 15-6
Items Deleted from the Reading Long-Term Trend Analysis

Age	Block	Item	Reason for Exclusion
9	BJ	N001801	Never scaled, extremely low probability of correct response
	BM	N003003	Excluded, poor fit to IRT model
	BJ	N008905	Excluded, marked score drift
13	BJ	N001801	Excluded, poor fit to IRT model
	BJ	N001904	Excluded, marked score drift
	BK	N002302	Never scaled, nonordinal item
	BL	N002804	Excluded, low interrater reliability
	BQ	N005001	Excluded, poor fit to IRT model
17	BJ	N001702	Nonmonotonic, was also dropped in 1992
	BK	N002302	Never scaled, nonordinal item
	BQ	N015905	Excluded, marked score drift

15.4 IRT SCALING FOR THE READING LONG-TERM TREND ASSESSMENT

15.4.1 Item Parameter Estimation

The first step in the scaling process was the estimation of item parameters for the long-term trend items. This item calibration was performed using the BILOG/PARSCALE program described in Chapter 11. Items were calibrated separately for each of the three age/grade groups. Item parameters were estimated using combined data from the assessment years 1992 and 1994, treating each assessment as a sample from a separate subpopulation. Student weights were used for the calibration. To ensure that each assessment year had a similar influence on the calibration, student weights for 1992 examinees were multiplied by a constant, to adjust them to have the same sum as the sum of the weights for the 1994 examinees. Approximately 600-700 examinees were present in each assessment year for each item.

Starting values for item parameters were based on the final item parameter values from the analysis of the 1992 long-term trend assessment. As described in Chapter 9, BILOG/PARSCALE calibrations were done in two stages. In stage one, the proficiency distribution of each assessment year was constrained to be normally distributed, although the means and variances differed across assessment years. The values of the item parameters from this normal solution were then used as starting values for a second-stage estimation run in which the proficiency distribution (modeled as a separate multinomial distribution for each assessment year) was estimated concurrently with item parameters. Calibration was concluded when changes in item parameters became negligibly small (i.e., less than .005). A list of the items scaled for each of the ages, along with their item parameter estimates, appears in Appendix D.

15.4.2 Evaluation of Model Fit

During and subsequent to item parameter estimation, evaluations of the fit of the IRT models were carried out for each of the items. These evaluations were based primarily on graphical analysis. First, model fit was evaluated by examining plots of nonmodel-based estimates of the expected proportion correct (conditional on proficiency) versus the proportion correct predicted by the estimated item response function (see Chapter 9 and Mislevy & Sheehan, 1987, p. 302). In making decisions about excluding items from the final scales, a balance was sought between being too stringent, hence deleting too many items and possibly damaging the content representativeness of the pool of scaled items, and being too lenient, hence including items with model fit poor enough to endanger the types of model-based inferences made from NAEP results. For the majority of the items, the model fit was extremely good.

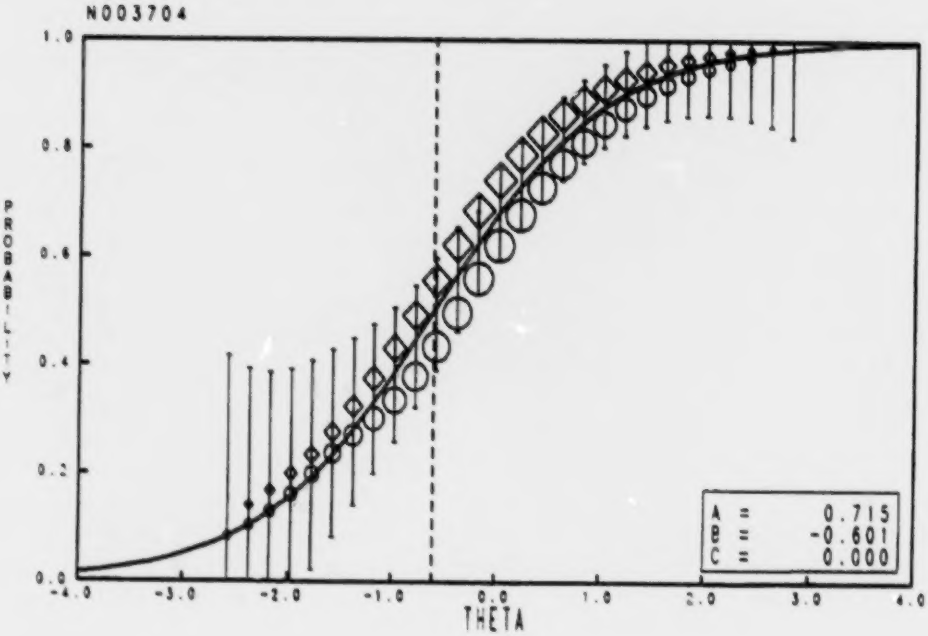
In addition to the constructed-response items dropped due to poor interrater reliability and item drift, one item from the age 9 long-term trend data and two items from the age 13 long-term trend data were deleted from scaling due to poor fit to the IRT model. The items were too difficult for these students to yield reliable estimates of item parameters. No other long-term trend items were deleted from the 1994 trend analysis. Table 15-6 lists items that were excluded from the long-term trend assessment. These items will be reexamined in future trend assessments.

The adequacy of the assumption of a common item response function across assessment years was also evaluated by comparing the nonmodel-based expected proportions for each assessment year to the single, model-based item response function fit by BILOG/PARSCALE. Items that showed clear evidence of functioning differently across assessments were treated as separate items for each assessment year—that is, separate item response functions were estimated for each assessment. As was the case with deleting items, in making decisions about scaling items separately by assessment year, a balance was sought between being too stringent, hence splitting too many items and possibly damaging the common item link between the assessment years, and being too lenient, hence including items with model fit poor enough to endanger the model-based trend inferences. These separately scaled items will be reexamined in future long-term trend assessments.

At each age, several items were calibrated separately for each assessment year because examination of residual plots identified some items as functioning differently across assessments. Figure 15-1 shows item N003704 from the analysis for grade 11/age 17. Data are presented for 1992 (diamonds) and for 1994 (ovals). For proficiency values less than 0.5, the two sets of symbols diverge, and the discrepancy is substantial for $\theta < 0.0$. The top (1992 data) and bottom (1994 data) of Figure 15-2 show the plots for the item treated separately by assessment year. The remaining misfit is relatively small, and affects a small proportion of the population. Overall, three of the 195 long-term trend reading items were calibrated separately by assessment year. Table 15-7 lists the items that were calibrated separately across assessment years.

Figure 15-1

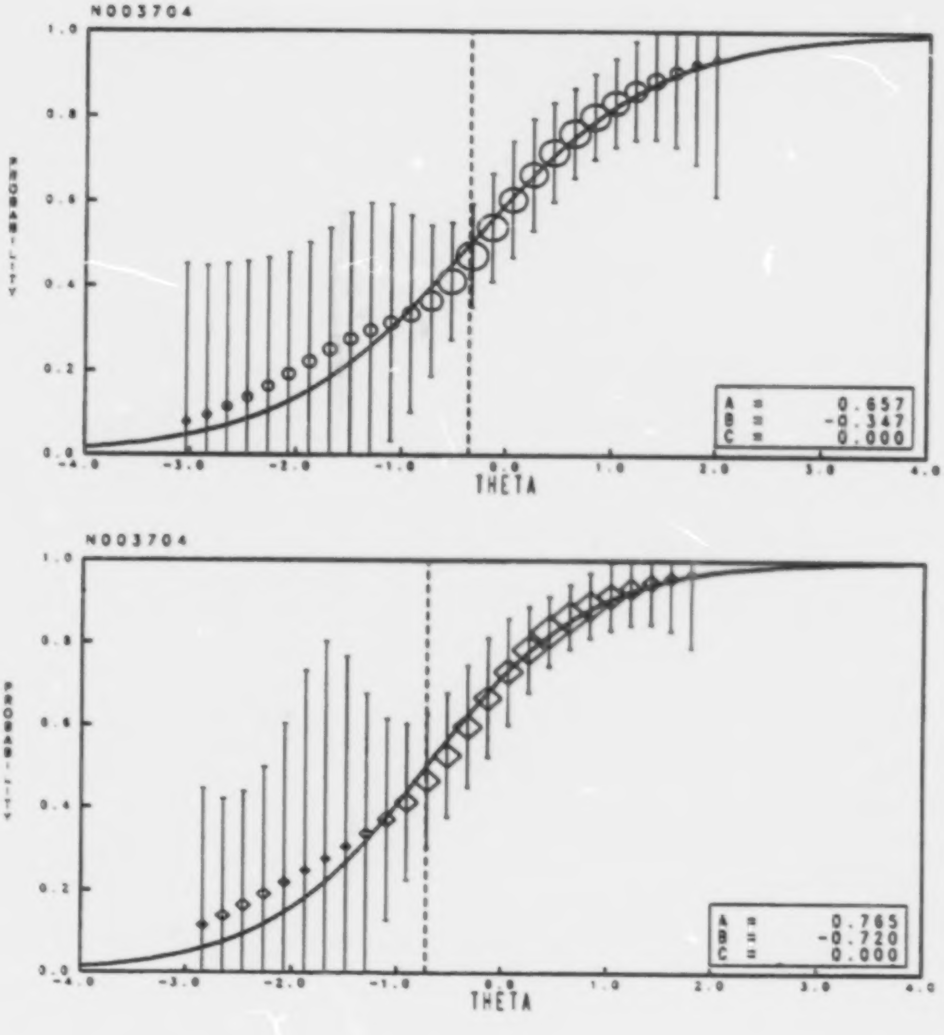
Example Long-Term Trend Item (N003704, Age 17)
*Demonstrating Differential Item Functioning Across Assessment Years 1992 and 1994**



****Note:** The plot compares empirical and model-based estimates of the item response function (IRF). The smooth curve represents the model-based estimate at each provisional proficiency level. The diamonds represent 1992 data; ovals represent 1994 data.*

Figure 15-2

Example Long-Term Trend Item (N003704, Age 17)
*Fitting Separate Item Response Functions for Each Assessment Year**



**__Note__: The plot compares empirical and model-based estimates of the item response function (IRF). The smooth curve represents the model-based estimate at each provisional proficiency level. The*

BEST COPY AVAILABLE

Table 15-7

Items Calibrated Separately by Assessment Year in the Reading Long-Term Trend Analysis

Age	Block	Item	Reason for Separate Calibration
9	BN	N003704	Poor fit across assessments to common item response function
17	BN	N003704	Poor fit across assessments to common item response function
	BP	N004605	Poor fit across assessments to common item response function

At age 13, two items (N002201 and N002202) caused difficulty in scaling. In preliminary calibrations, both items did not fit the model well and had extremely large a-parameter values (4.5 and 6.3, respectively). The item response function of N002202 also demonstrated an elevated tail when compared to previous assessments. Examination of the content of the items indicated that this behavior might be due to local dependence of the items, although neither item had been problematic in previous assessments. A two-stage approach was used to yield stable item parameter estimates. First, item N002202 was temporarily deleted from scaling and all items were calibrated to convergence. In stage two, the item parameters of N002201 were fixed at their values from the stage-one calibration. Item N002201 was reintroduced into the scale and all item parameters were calibrated to convergence. This procedure yielded estimates that were similar to those obtained in previous assessments and exhibited acceptable model-data fit for all items. Item parameter estimates from stage two served as the final estimates for age 13.

15.5 GENERATION OF PLAUSIBLE VALUES

The generation of plausible values was conducted independently for each age/grade level for each of the assessment years. The item parameters from BILOG/PARSCALE, final student weights, item responses, and selected background variables were used with the computer program BGROUP (described in Chapter 11) to generate the values for each age. The background variables included student demographic characteristics (i.e., race/ethnicity of the student, highest level of education attained by parents), students' perceptions about reading, and student behavior both in and out of school (i.e., amount of television watched daily, amount of homework done each day). Appendix C gives the codings for the conditioning variables for the three age groups. Table 15-8 contains a list of the number of background contrasts included in conditioning, as well as the proportion of variance accounted for by the conditioning model for each age/grade.

Table 15-8

Proportion of Proficiency Variance Accounted for by the Conditioning Model for the Reading Long-Term Trend Assessment

Age/Grade	Number of Conditioning Contrasts*	Proportion of Proficiency Variance
9/4	49	.360
13/8	48	.388
17/11	47	.361

***Note**: *Excluding the constant term.*

15.6 THE FINAL READING LONG-TERM TREND SCALE

The linear indeterminacy of the long-term trend scale was resolved by linking the 1994 long-term trend scales to previous long-term trend scales. For each age, the item parameters from the joint calibration based on data from 1992 and 1994 were used with the 1992 data to find plausible values for the 1992 data. The mean and standard deviation of all of the plausible values were calculated and matched to the mean and standard deviation of all of the plausible values based on the original analysis of the 1992 data, as given in earlier reports. The transformations that resulted from this matching of the first two moments for the 1992 data are

$$\text{Age 9: } \theta_{\text{proficiency}} = 41.71 \bullet \theta_{\text{calibrated}} + 210.98,$$

$$\text{Age 13: } \theta_{\text{proficiency}} = 38.81 \bullet \theta_{\text{calibrated}} + 257.91, \text{ and}$$

$$\text{Age 17: } \theta_{\text{proficiency}} = 44.45 \bullet \theta_{\text{calibrated}} + 283.99,$$

where $\theta_{\text{proficiency}}$ denotes values on the final transformed scale and $\theta_{\text{calibrated}}$ denotes values on the calibration scale. Overall summary statistics for the long-term trend samples are given in Table 15-9.

As in the past, interpretation of the long-term trend results was facilitated through the provision of scale anchoring information. In 1984, five NAEP reading scale levels were selected as anchor points. These points (described in *Trends in Academic Progress*) are:

150 = simple, discrete reading tasks;
200 = partially developed skills and understanding;
250 = interrelation of ideas and generalizations;
300 = understanding complicated information; and
350 = learning from specialized reading materials.

Table 15-9

Means and Standard Deviations on the Reading Long-Term Trend Scale

Age	Assessment	All Five Plausible Values	
		Mean	S. D.
9	1984	211.0	41.1
	1988	211.8	41.2
	1990	209.2	44.7
	1992	210.5	40.4
	1994	211.0	40.5
13	1984	257.1	35.5
	1988	257.5	34.7
	1990	256.8	36.0
	1992	259.8	39.4
	1994	257.9	39.8
17	1984	288.8	40.3
	1988	290.1	37.1
	1990	290.2	41.3
	1992	289.7	43.0
	1994	288.1	44.4

Detailed descriptions of the skills required to read at each level were derived and benchmark exercises were selected to exemplify each level. These same anchor points were used in the 1988, 1990, 1992, and 1994 reading long-term trend reports. The estimated proportion of students in each reporting category who are at or above each anchor point were examined in the *NAEP 1994 Trends in Academic Progress: Achievement of U.S. Students in Science, 1969 to 1994; Mathematics, 1973 to 1994; Reading, 1971 to 1994; and Writing, 1984 to 1994* (Campbell, Reese, O'Sullivan, Dossey, & Donahue, 1996).

15.7 PARTITIONING OF THE ESTIMATION ERROR VARIANCE

For each age, the variance of the final, transformed scale mean was partitioned as described in Chapter 11. This analysis yielded estimates of the proportion of error variance due to sampling students and the proportion due to the latent nature of θ. These estimates are given in Table 15-10 (for stability of the estimates, they are based on 100 plausible values). More detailed information is available for gender and race/ethnicity subgroups in Appendix E.

Table 15-10

Estimation Error Variance and Related Coefficients for the Reading Long-Term Trend Assessment

Age	Total Estimation of Error Variance	Proportion of Variance Due to ...	
		Student Sampling	Latency of θ
9	1.55	.88	.12
13	.85	.83	.17
17	2.02	.86	.14

Chapter 16

DATA ANALYSIS FOR THE LONG-TERM TREND MATHEMATICS ASSESSMENT[1]

Edward Ip, Frank Jenkins, Edward Kulick, Hong Zhou, and Stephen Szyszkiewicz
Educational Testing Service

16.1 INTRODUCTION

This chapter describes the analyses performed on the responses to the cognitive and background items in the 1994 assessment of mathematics. These analyses led to the results presented in the *NAEP 1994 Trends in Academic Progress: Achievement of U.S. Students in Science, 1969 to 1994; Mathematics, 1973 to 1994; Reading, 1971 to 1994; and Writing, 1984 to 1994* (Campbell, Reese, O'Sullivan, Dossey, & Donahue, 1996). The emphasis of this chapter is on the methods and results of procedures used to develop the IRT-based scale scores; however, some attention is given to the analysis of constructed-response items. The theoretical underpinnings of the IRT and plausible values methodology described in this chapter are given in Chapter 11.

The objectives of the mathematics analyses were to prepare scale values and perform all analyses necessary to produce a long-term trend report in mathematics. The mathematics long-term trend results include the years 1973, 1978, 1982, 1986, 1990, 1992, and 1994.

The student samples that were administered mathematics items in the 1994 long-term trend assessment are shown in Table 16-1. See Chapters 1 and 3 for descriptions of the target populations and the sample design used for the assessment.

---

[1] Edward Ip was the primary person responsible for the planning, specification, and coordination of the mathematics long-term trend analyses. Frank Jenkins advised and reviewed the scaling of the long-term trend mathematics data. Computer activities for all long-term trend mathematics scaling and data analyses were directed by Edward Kulick and performed by Hong Zhou and Stephen Szyszkiewicz.

414

Table 16-1

NAEP 1994 Mathematics Student Samples

Sample	Booklets	Mode	Cohort Assessed	Time of Testing	Age Definition	Modal Grade	Number Assessed
9 [MS-LTTrend]	91-93	Tape	Age 9	Winter	CY	4	5,663
13 [MS-LTTrend]	91-93	Tape	Age 13	Fall	CY	8	6,052
17 [MS-LTTrend]	84-85	Tape	Age 17	Spring	Not CY	11	5,702

LEGEND

MS	*Mathematics and science*
LTTrend	*Long-term trend: booklets are identical to long-term trend assessment of 1986*
Tape	*Audiotape administration*
CY	*Calendar year: birthdates in 1984, 1980, and 1976 for ages 9, 13, and 17*
Not CY	*Age 17 only: birthdates between October 1, 1976, and September 30, 1977*

The long-term trend results reported in the *NAEP 1994 Trends in Academic Progress: Achievement of U.S. Students in Science, 1969 to 1994; Mathematics, 1973 to 1994; Reading, 1971 to 1994; and Writing, 1984 to 1994* are based on paced-tape administrations and occur at all age levels. The samples involved in the analysis are shown as 9[MS-LTTrend], 13[MS-LTTrend], and 17[MS-LTTrend] in Table 16-1. For ages 9 and 13, the long-term trend booklets contained blocks of mathematics and science items administered by audiotape (these booklets also contained reading blocks administered in print form). The age 17 long-term trend booklets contained only mathematics and science blocks, both presented by paced-tape recordings. All students received a block of common background questions, distinct for each age. Subject-area background questions were presented in the cognitive blocks. The booklets for the age 9 and age 13 samples (booklets 91-93) are the same as those used for long-term trend assessments in 1986, 1988, 1990, and 1992. The booklets for the age 17 sample (booklets 84-85) are the same as those used for the 1986, 1990, and 1992 long-term trend assessments. The booklets and the blocks within those booklets are listed in Tables 4-20 through 4-22 in Chapter 4.

Table 16-2 clarifies the relationships among the 1994 long-term trend samples and samples from previous years. For ages 9, 13, and 17, the paced-tape bridge to the 1986 samples allows direct comparisons with 1992 samples and with 1986 long-term trend samples. There was also a paced-tape administration in 1988, at ages 9 and 13, that was comparable to the other years. However, a paced-tape administration was not conducted at age 17 in 1988. Instead, a noncomparable paper-based assessment was administered. Hence, 1988 is not included as a point in the long-term trend reporting. In 1986, the mathematics long-term trend items were scaled with common items from the 1978 and 1982 assessments. Because the 1973 assessment had few items in common with the current assessment, data from that assessment was not scaled using the IRT model but was linked to the trend line by a linear transformation involving the mean proportion correct for common items. The 1994 long-term trend assessment was linked to the 1973, 1978, and 1982 assessments through the 1986 assessment. Information about previous assessment years is available in *Expanding the New Design: The NAEP 1985-86 Technical Report* (Beaton, 1988) and *The NAEP 1992 Technical Report* (Johnson & Carlson, 1994).

Table 16-2

NAEP Mathematics Samples Contributing to 1994 Long-Term Trend Results, 1973-1994

Cohort Assessed	Year	Sample	Subjects	Time of Testing	Mode of Administration	Age Definition	Modal Grade
Age 9	1973	Main	MS	Winter	Tape	CY	4
	1978	Main	M	Winter	Tape	CY	4
	1982	Main	MCS	Winter	Tape	CY	4
	1986	LTTrend*	MS	Winter	Tape**	CY	4
	1990	LTTrend*	MS	Winter	Tape**	CY	4
	1992	LTTrend*	MS	Winter	Tape	CY	4
	1994	LTTrend*	MS	Winter	Tape	CY	4
Age 13	1973	Main	MS	Fall	Tape	CY	8
	1978	Main	M	Fall	Tape	CY	8
	1982	Main	MCS	Fall	Tape	CY	8
	1986	LTTrend*	MS	Fall	Tape**	CY	8
	1990	LTTrend*	MS	Fall	Tape**	CY	8
	1992	LTTrend*	MS	Fall	Tape	CY	8
	1994	LTTrend*	MS	Fall	Tape	CY	8
Age 17	1973	Main	MS	Spring	Tape	Not CY	11
	1978	Main	M	Spring	Tape	Not CY	11
	1982	Main	MCS	Spring	Tape	Not CY	11
	1986	LTTrend*	MS	Spring	Tape	Not CY	11
	1990	LTTrend*	MS	Spring	Tape	Not CY	11
	1992	LTTrend*	MS	Spring	Tape	Not CY	11
	1994	LTTrend*	MS	Spring	Tape	Not CY	11

* ***Note:*** *Within an age group, these samples received common booklets.*

** ***Note:*** *Mathematics and science administered by audiotape, reading administered by print.*

LEGEND

M	*Mathematics*	*Main*	*Main Assessment*
MS	*Mathematics and science*	*Tape*	*Audiotape administration*
MCS	*Mathematics, civics, and science*	*CY*	*Calendar year: birthdates in 1984, 1980, and 1976 for ages 9, 13, and 17*
LTTrend	*Long-term trend: booklets are identical to the long-term trend assessment of 1986.*	*Not CY*	*Age 17 only: birthdates between October 1, 1976, and September 30, 1977*

The steps in the mathematics long-term trend analysis are documented in the following sections. Consistent with procedures followed in earlier NAEP analyses, the first step was to gather item-level and block-level information. Next, the items were calibrated and evaluated for IRT model fit. Derived background variables were calculated for use in conditioning and reporting. Item parameter estimates were then used to generate plausible values after conditioning on available background variables. The scales were then placed on the mathematics long-term trend scale used in previous trend assessments.

Table 16-3 indicates the number of items in common across different age combinations. Table 16-4 shows the number of items scaled in 1994 that were common across assessment years. The 1986, 1990, 1992, and 1994 assessments had all items in common. For age 9, the number of items common across assessment years 1978 to 1994 was only 35. For age 13, the overlap across all assessments was 56 items; for age 17, the overlap was 54 items. Item parameters were estimated assuming a univariate scale, since the number of items presented to each student was small and there were too few items to estimate several content area scales separately.

Table 16-3
Numbers of Scaled Mathematics Long-Term Trend Items Common Across Ages

Age	Booklets	Number of Items
9 only	91-93	32
13 only	91-93	30
17 only	84-85	41
9 and 13 only	91-93, 91-93	20
9 and 17 only	91-93, 84-85	0
13 and 17 only	91-93, 84-85	27
9, 13, and 17	91-93, 91-93, 84-85	3
Total		**153**

Table 16-4
Numbers of Scaled Mathematics Long-Term Trend Items Common Across Assessments

	Number of Items		
Assessment Year	Age 9	Age 13	Age 17
1986, 1990, 1992, 1994	55	80	71
1982, 1986, 1990, 1992, 1994	53	79	65
1978, 1986, 1990, 1992, 1994	35	56	54
1978, 1982, 1986, 1990, 1992, 1994	35	56	54

In the first phase of the analysis, standard item statistics were calculated. The results served as a check for data entry errors and as a reasonableness check against results from previous assessments.

In the second phase, the IRT model was fitted to the data across multiple assessments for each age separately. This procedure puts item parameters and ability estimates on the same scale across years. The same item may have different item parameters for different age groups.

The analysis for an age group was completed by the creation of plausible values through a multiple imputation estimation procedure in which item parameter estimates, student responses, and student background information were combined to produce the most precise possible estimates of student ability. Plausible values from the 1994 assessment were transformed to the scale of the 1986 proficiency measures.

The 1994 plausible values for ages 9, 13, and 17 were used to create proficiency means and jackknife estimates of standard errors for the whole group and for subgroups. These proficiency means form the final point in the mathematics long-term trend from 1973 to 1994.

The specifics of the mathematics long-term trend analysis are documented in the following sections.

16.2 ITEM ANALYSIS FOR THE MATHEMATICS LONG-TERM TREND ASSESSMENT

A coding problem on item N260601 was detected in the early stage of analysis. The item was rescored by NCS. No other problems in coding, formats, or data were detected. The conventional item analysis, with results displayed in Table 16-5, was performed at the block level on the paced-tape long-term trend data.

Table 16-5 contains the number of items, size of the sample administered to the block, mean weighted proportion correct, mean weighted r-biserial, and mean weighted alpha as a measure of reliability for each block. The average values were calculated using examinee weights and the items in the block that were scaled. The 1994 item-level statistics were not very different from those for the 1992 assessment. Similar statistics for the 1992 assessment were reported in Table 13-6 of the 1992 Technical Report. Because errors were found in that table, an updated version is provided in Table 16-6.

The item-level omit rates show, on average, a slight but consistent increase from 1992 to 1994. For age 9, the average omit rate across blocks increases from 2.0 percent (with a standard error of 1.9) to 2.2 percent (with a standard error of 1.8). For age 13, the average omit rate across blocks increases from 1.1 percent (1.5) to 1.3 percent (1.2). For age 17, the increase in average omit rate across blocks is more noticeable. It increases from 1.6 percent (2.1) to 2.3 percent (2.6). Moreover, for age 17, the omit rate increased in 1994 for all but three items (of the total of 71). The percent of examinees not reaching items in the long-term trend blocks was almost always zero because the items were administered with a tape recording to pace response time.

The correspondence between blocks, booklets, and samples is given for the mathematics long-term trend assessment in Tables 4-20 through 4-22 in Chapter 4. Common labeling of these blocks across ages does not denote common items.

418

Table 16-5

Descriptive Statistics for Item Blocks in the Mathematics Long-Term Trend Samples (1994)

AGE 9

Statistic	Blocks		
	M1	M2	M3*
Number of scaled items	24	26	5
Number of scaled constructed-response items	9	9	0
Unweighted sample size	1,892	1,893	1,878
Average weighted proportion correct	.63	.63	.68
Average weighted r-biserial	.61	.63	.82
Weighted alpha reliability	.82	.85	.49

AGE 13

Statistic	Blocks		
	M1	M2	M3*
Number of scaled items	36	36	8
Number of scaled constructed-response items	9	8	0
Unweighted sample size	2,040	2,059	1,953
Average weighted proportion correct	.67	.63	.65
Average weighted r-biserial	.60	.56	.68
Weighted alpha reliability	.87	.85	.61

AGE 17

Statistic	Blocks		
	M1	M2	M3*
Number of scaled items	33	33	5
Number of scaled constructed-response items	10	5	1
Unweighted sample size	1,889	1,889	1,924
Average weighted proportion correct	.64	.64	.57
Average weighted r-biserial	.67	.62	.78
Weighted alpha reliability	.90	.88	.56

****Note*:** *This block is mostly calculator items, which were not analyzed. For the item analysis, students who did not respond to any items in the block were omitted; however, such students were assigned proficiencies in the final database.*

Table 16-6

*Descriptive Statistics for Item Blocks in the Mathematics Long-Term Trend Samples (1992)**

AGE 9

Statistic	Blocks		
	M1	**M2**	**M3****
Number of scaled items	24	26	5
Number of scaled constructed-response items	9	9	0
Unweighted sample size	2,338	2,435	2,512
Average weighted proportion correct	.61	.63	.69
Average weighted r-biserial	.61	.63	.82
Weighted alpha reliability	.82	.86	.46

AGE 13

Statistic	Blocks		
	M1	**M2**	**M3****
Number of scaled items	36	36	8
Number of scaled constructed-response items	9	8	0
Unweighted sample size	1,928	2,005	1,976
Average weighted proportion correct	.67	.61	.66
Average weighted r-biserial	.57	.56	.67
Weighted alpha reliability	.86	.85	.60

AGE 17

Statistic	Blocks		
	M1	**M2**	**M3****
Number of scaled items	33	33	5
Number of scaled constructed-response items	10	5	1
Unweighted sample size	2,207	2,207	2,152
Average weighted proportion correct	.65	.66	.57
Average weighted r-biserial	.67	.63	.75
Weighted alpha reliability	.90	.88	.51

**Note*: This table has been updated since its publication in The NAEP 1992 Technical Report (Table 13-6).*

***Note*: This block is mostly calculator items, which were not analyzed. For the item analysis, students who did not respond to any items in the block were omitted; however, such students were assigned proficiencies in the final database.*

BEST COPY AVAILABLE

16.3 IRT SCALING FOR THE MATHEMATICS LONG-TERM TREND ASSESSMENT

16.3.1 Item Parameter Estimation

The scaling process began with the estimation of item parameters. IRT parameters were estimated using the NAEP version of the BILOG/PARSCALE program described in Chapter 11. Item calibration was performed separately for each of the three age groups, using the total combined data from the 1992 and 1994 assessments. Including the 1992 assessment data assures that item parameters will be similar for adjacent assessments so that year-to-year trends will not be distorted by abrupt changes in calibration. The calibration was performed on the entire sample of students, resulting in a range of about 1,800 to 2,600 examinee responses to each item in each assessment year. The calibration was based on student weights that were rescaled for the 1994 data so that the sum of the weights equaled the unweighted sample size. Also, weights for the 1994 data were restandardized to give equal weight to the two assessment years included in the scaling. As with the previous assessment, calculator items were excluded from the analysis. Because calculators have changed greatly since the start of the long-term trend assessment, it was judged that calculator questions are no longer comparable across time. These items were kept in the assessment, since excluding them would have changed the testing context.

Since parameters for items in blocks M1, M2, and M3 were estimated separately for ages 9, 13, and 17, items administered at more than one age have multiple sets of item parameter estimates. Items were examined for lack of fit with the data. Those that exhibited extreme violation of IRT assumptions (i.e., did not have monotonically increasing item characteristic curves) were deleted from the analysis, as they were in previous assessments. Other items were deleted because they were calculator items, which were not considered part of the regular assessment. These excluded items appear in Tables 16-7, 16-8, and 16-9. As a result of these deletions, 55 items were scaled for age 9, 80 items were scaled for age 13, and 71 items were scaled for age 17. Of the 153 noncalculator items that were part of the assessment, seven items (5%) were excluded due to poor fit with the data. A list of the items scaled for each of the ages, along with their item parameter estimates, appears in Appendix D.

Table 16-7
Items Deleted from the Age 9 Mathematics Long-Term Trend Analysis

Booklet	Block	Item	Reason for Exclusion
91	M1	N252601	Was deleted in prior assessment
		N262502	Was deleted in prior assessment
92	M3	N268221	Calculator item*
		N276021	Calculator item
		N276022	Calculator item
		N276821	Calculator item
		N276822	Calculator item
		N276823	Calculator item
		N277621	Calculator item
		N277622	Calculator item
		N277623	Calculator item
		N284021	Calculator item
		N284022	Calculator item

**Note: All calculator items were deleted from the analysis.*

Table 16-8

Items Deleted from the Age 13 Mathematics Long-Term Trend Analysis

Booklet	Block	Item	Reason for Exclusion
91	M1	N262502	Was deleted in prior assessment
93	M2	N261601	Was deleted in prior assessment
92	M3	N264521	Calculator item*
		N259921	Calculator item
		N276821	Calculator item
		N276822	Calculator item
		N276823	Calculator item
		N278921	Calculator item
		N278922	Calculator item
		N278923	Calculator item
		N278924	Calculator item
		N278925	Calculator item
		N280621	Calculator item
		N280622	Calculator item
		N280623	Calculator item
		N280624	Calculator item
		N280625	Calculator item
		N280626	Calculator item

__Note__: All calculator items were deleted from the analysis.

Table 16-9

Items Deleted from the Age 17 Mathematics Long-Term Trend Analysis

Booklet	Block	Item	Reason for Exclusion
84	M1	N282801	Was deleted in prior assessment
		N285701	Was deleted in prior assessment
84	M2	N266801	Was deleted in prior assessment
		N255301	Was deleted in prior assessment
85	M3	N259921	Calculator item*
		N264321	Calculator item
		N264521	Calculator item
		N267921	Calculator item
		N276821	Calculator item
		N276822	Calculator item
		N276823	Calculator item
		N278921	Calculator item
		N278922	Calculator item
		N278923	Calculator item
		N278924	Calculator item
		N278925	Calculator item
		N280621	Calculator item
		N280622	Calculator item
		N280623	Calculator item
		N280624	Calculator item
		N280625	Calculator item
		N280626	Calculator item
		N285321	Calculator item

***Note:** All calculator items were deleted from the analysis.*

16.4 DERIVED BACKGROUND VARIABLES

In the long-term trend analysis, all derived variables based upon background questions were used for conditioning and in reporting (to define subgroups). Derived reporting and conditioning variables are described in Appendix B.

16.5 GENERATION OF PLAUSIBLE VALUES

The generation of plausible values was conducted independently for each age group. In this approach, student background information was used to condition item responses in order to more accurately estimate student abilities. The univariate conditioning program BGROUP was used to combine NAEP BILOG/PARSCALE item parameters with weighted item responses and background variables to produce posterior ability estimates called plausible values. As defined in Chapter 11, BGROUP is an enhanced version of the original conditioning program, MGROUP. *Plausible values are not test scores* in the usual sense, but can be used to provide consistent estimates of population characteristics. There were 53 contrasts in the conditioning model at age 9, 56 at age 13, and 63 at age 17. Appendix C gives the codings for the conditioning variables for the three age groups. The estimated conditioning effects in the tables are expressed on the scale of the original calibration (i.e., the theta scale). A check was made on the distributions of the plausible values for each age, including inspection of the whole group and subgroup means and standard deviations. Table 16-10 contains a list of the number of background contrasts included in conditioning, as well as the proportion of variance accounted for by the conditioning model for each age/grade.

Table 16-10
Proportion of Proficiency Variance Accounted for by the Conditioning Model for the Mathematics Long-Term Trend Assessment

Age/Grade	Number of Conditioning Contrasts*	Proportion of Proficiency Variance
9/4	53	.405
13/8	56	.394
17/12	63	.592

***Note**: *Excluding the constant term.*

16.6 THE FINAL MATHEMATICS LONG-TERM TREND SCALE

Since the plausible value (theta) scales have a linear indeterminacy, comparisons with previous assessments will be sensible only if the scale is linearly transformed to a meaningful metric. This indeterminacy was resolved by linking the 1994 scales to previous long-term trend scales. The 1994 data had to be transformed to compensate for linear changes in the scale due to employing newly estimated item parameters and new BGROUP conditioning parameters in 1994. The transformation was accomplished by first reestimating the 1992 student abilities using 1994 item parameters and 1994 BGROUP parameters. The new 1992 ability estimates were then equated to the old 1992 ability estimates by matching the first two moments (i.e., the mean and standard deviation). The constants for this transformation were then applied to the 1994 data. The transformation equations that resulted are

$$\text{Age 9:}\ \ \theta_{\text{proficiency}} = 33.66 \bullet \theta_{\text{calibrated}} + 229.73,$$

$$\text{Age 13:}\ \ \theta_{\text{proficiency}} = 33.25 \bullet \theta_{\text{calibrated}} + 273.29, \text{ and}$$

$$\text{Age 17:}\ \ \theta_{\text{proficiency}} = 30.39 \bullet \theta_{\text{calibrated}} + 306.07,$$

where $\theta_{\text{proficiency}}$ denotes an individual's value on the final transformed scale of the 1994 data and $\theta_{\text{calibrated}}$ denotes an individual's value on the original 1994 theta scale. Overall summary statistics for the long-term trend samples are given in Table 16-11.

Table 16-11

Means and Standard Deviations on the Mathematics Long-Term Trend Proficiency Scale

Age	Assessment	All Five Plausible Values	
		Mean	**S. D.**
9	1978	218.6	36.0
	1982	219.0	34.8
	1986	221.7	34.0
	1990	229.6	32.9
	1992	229.6	33.1
	1994	231.1	33.2
13	1978	264.1	39.0
	1982	268.6	33.4
	1986	269.0	30.8
	1990	270.4	31.3
	1992	273.1	30.9
	1994	274.3	32.4
17	1978	300.4	34.9
	1982	298.5	32.4
	1986	302.0.	31.0
	1990	304.6	31.3
	1992	306.7	30.1
	1994	306.2	30.2

To provide a context for interpreting the overall mathematics long-term trend results, the NAEP mathematics results were "anchored" at five NAEP mathematic scale levels. These points (described in the *NAEP 1994 Trends in Academic Progress: Achievement of U.S. Students in Science, 1969 to 1994; Mathematics, 1973 to 1994; Reading, 1971 to 1994; and Writing, 1984 to 1994*) are:

150 = simple arithmetic facts;
200 = beginning skills and understanding;
250 = numerical operations and beginning problem solving;
300 = moderately complex procedures and reasoning; and
350 = multi-step problem solving and algebra.

These same anchor points were used in the 1978, 1982, 1986, 1990, and 1992 reading long-term trend reports.

16.7 PARTITIONING OF THE ESTIMATION ERROR VARIANCE

For each age's scale, the error variance of the final, transformed proficiency mean was partitioned as described in Chapter 11. The variance is partitioned into two parts; the proportion of error variance due to sampling students (sampling variance) and the proportion of error variance due to the fact that proficiency, θ, is a latent variable that is estimated rather than observed. Table 16-12 contains estimates of the total error variance, the proportion of error variance due to sampling students, and the proportion of error variance due to the latent nature of θ (for stability of the estimates, these are based on 100 plausible values). More detailed information is available for gender and race/ethnicity subgroups in Appendix E.

Table 16-12
Estimation Error Variance and Related Coefficients
for the Mathematics Long-Term Trend Assessment

Age	Total Estimation Error Variance	Proportion of Variance Due to ...	
		Student Sampling	Latency of θ
9	0.62	0.87	0.13
13	0.94	0.95	0.06
17	0.91	0.93	0.07

Chapter 17

DATA ANALYSIS FOR THE LONG-TERM TREND SCIENCE ASSESSMENT[1]

Spencer S. Swinton, Nancy L. Allen, Steven P. Isham, and C. Janet Chen
Educational Testing Service

17.1 INTRODUCTION

This chapter describes the analyses performed on the responses to the cognitive and background items in the 1994 long-term trend assessment of science. These analyses led to the results presented in the *NAEP 1994 Trends in Academic Progress: Achievement of U.S. Students in Science, 1969 to 1994; Mathematics, 1973 to 1994; Reading, 1971 to 1994; and Writing, 1984 to 1994* (Campbell, Reese, O'Sullivan, Dossey, & Donahue, 1996). The emphasis of this chapter is on the methods and results of procedures used to develop the IRT-based scale scores that formed the basis of these reports. The theoretical underpinnings of the IRT and plausible value methodology described in this chapter are provided in Chapter 11.

The objective of the science analyses was to prepare scale values and perform all analyses necessary to produce a long-term trend report in science. The science long-term trend line includes the years 1970, 1973, 1977, 1982, 1986, 1990, 1992, and 1994.

The student samples that were administered science items in the 1994 assessment are shown in Table 17-1. (See Chapters 1 and 3 for descriptions of the target populations and the sample design used for the assessment.)

---

[1] Spencer S. Swinton was the primary person responsible for the planning, specification, and coordination of the science long-term trend analyses. Nancy L. Allen advised and reviewed the scaling of the long-term trend science data. Computer activities for all long-term trend science scaling and data analyses were directed by Steven P. Isham and performed by C. Janet Chen.

Table 17-1

NAEP 1994 Science Student Samples

Sample	Booklets	Mode	Cohort Assessed	Time of Testing	Age Definition	Modal Grade	Number Assessed
9 [MS-LTTrend]	91-93	Tape	Age 9	Winter	CY	4	5,663
13 [MS-LTTrend]	91-93	Tape	Age 13	Fall	CY	8	6,052
17 [MS-LTTrend]	84-85	Tape	Age 17	Spring	Not CY	11	3,813

LEGEND

MS	*Mathematics and science*
LTTrend	*Long-term trend*
Tape	*Audiotape administration*
CY	*Calendar year: birthdates in 1984, 1980, and 1976 for ages 9, 13, and 17*
Not CY	*Age 17 only: birthdates between October 1, 1976, and September 30, 1977*

The science long-term trend results reported in the *NAEP 1994 Trends in Academic Progress: Achievement of U.S. Students in Science, 1969 to 1994; Mathematics, 1973 to 1994; Reading, 1971 to 1994; and Writing, 1984 to 1994* are based on paced-tape administrations at all three age levels. The samples involved in the analysis were samples 9[MS-LTTrend], 13[MS-LTTrend], and 17[MS-LTTrend] in Table 17-1. For ages 9 and 13, the long-term trend booklets for these samples contained blocks of reading, mathematics, and science items. The science and mathematics blocks were paced by tape-recordings and the reading blocks were presented in print form. The age 17 long-term trend booklets contained only mathematics and science blocks, both presented by paced-tape recordings. All students received a block of common background questions, distinct for each age. Subject-area background questions were presented in the cognitive blocks. The booklets for the age 9 and age 13 samples (booklets 91-93) and the booklets for the age 17 samples (booklets 84-85) are the same as those used for long-term trend assessments in 1986, 1990, and 1992. The booklets and the blocks within those booklets are listed in Chapter 4. Additional information about all of the items in these blocks is in also found in that chapter. This chapter includes specific information about the long-term trend items that were scaled.

Table 17-2 clarifies the relationships between the 1994 long-term trend samples and samples from previous years. For all ages, the paced-tape bridge from the 1994 assessment to the 1986 assessment allows direct comparisons with 1990, 1992, and 1994 samples and with 1986 long-term trend and main samples. There was also a tape administration in 1988 at ages 9 and 13 that was comparable to the other years. However, a tape administration was not conducted at age 17 in 1988. Instead, a noncomparable paper-based assessment was conducted. Hence, 1988 is not included as a point in the long-term trend reporting. In 1986, the science long-term trend items were scaled with common items from the 1977 and 1982 assessments. Because of the small number of items in common with those in the 1970 and 1973 assessments, data from those assessments were not scaled, but were linked to the long-term trend line through mean proportion correct for common items. The 1990 long-term trend assessments were linked to the 1970, 1973, 1977, and 1982 assessments through the 1986 assessment. The 1994 long-term trend assessments were linked to previous assessments through the 1992 long-term trend assessment, and thence through the 1990 assessment. Information about previous assessment years, including 1970 and 1973, is available in Chapter 11 of *Expanding the New Design: The NAEP 1985-86 Technical Report* (Yamamoto, 1988), Chapter 14 of *The NAEP 1990 Technical Report* (Allen, 1992), and Chapter 14 of *The NAEP 1992 Technical Report* (Allen & Isham, 1994).

The numbers of scaled items for each age are presented in Table 17-3. As was done with previous long-term trend analyses, each age was scaled separately and the long-term trend scales are univariate. Derivation of scales for specific content areas was not feasible given the limited number of items presented to students in the long-term trend samples. The number of items scaled in 1994 that were common across assessment years is presented in Table 17-4.

The steps in the science long-term trend analysis are documented in the following sections. As is usual in NAEP analyses, the first step was to gather item and block information. Next, the long-term trend items were calibrated, derived background variables were calculated, and plausible values were generated after conditioning on available background variables. Finally, the scales were placed on the final science long-term trend scale used in previous trend assessments.

Table 17-2

NAEP Science Samples Contributing to 1994 Long-Term Trend Results, 1970-1994

Cohort Assessed	Year	Sample	Subjects	Time of Testing	Mode of Administration	Age Definition	Modal Grade
Age 9	1970	Main	SWC	Winter	Tape	CY	4
	1973	Main	MS	Winter	Tape	CY	4
	1977	Main	SCI	Winter	Tape	CY	4
	1982	LTTrend*	MSC	Winter	Tape	CY	4
	1986	LTTrend*	MS	Winter	Tape**	CY	4
	1990	LTTrend*	MS	Winter	Tape**	CY	4
	1992	LTTrend*	MS	Winter	Tape**	CY	4
	1994	LTTrend*	MS	Winter	Tape**	CY	4
Age 13	1970	Main	SWC	Fall	Tape	CY	8
	1973	Main	MS	Fall	Tape	CY	8
	1977	Main	SCI	Fall	Tape	CY	8
	1982	LTTrend*	MSC	Fall	Tape	CY	8
	1986	LTTrend*	MS	Fall	Tape**	CY	8
	1990	LTTrend*	MS	Fall	Tape**	CY	8
	1992	LTTrend*	MS	Fall	Tape**	CY	8
	1994	LTTrend*	MS	Fall	Tape**	CY	8
Age 17	1970	Main	SWC	Spring	Tape	Not CY	11
	1973	Main	MS	Spring	Tape	Not CY	11
	1977	Main	SCI	Spring	Tape	Not CY	11
	1982	LTTrend*	MSC	Spring	Tape	Not CY	11
	1986	LTTrend*	MS	Spring	Tape	Not CY	11
	1990	LTTrend*	MS	Spring	Tape	Not CY	11
	1992	LTTrend*	MS	Spring	Tape	Not CY	11
	1994	LTTrend*	MS	Spring	Tape	Not CY	11

* ***Note:*** *Within an age group, these samples received common booklets.*

*****Note:*** *Mathematics and science administered by audiotape, reading administered by print.*

LEGEND

SCI	*Science*	*LTTrend*	*Long-term trend: booklets are identical to the long-term trend assessment of 1986*
MS	*Mathematics and science*		
MSC	*Mathematics, science, and citizenship*	*Tape*	*Audiotape administration*
SWC	*Science, writing, and citizenship*	*CY*	*Calendar year: birthdates in 1984, 1980, and 1976 for ages 9, 13, and 17*
Main	*Main assessment*	*Not CY*	*Age 17 only: birthdates between October 1, 1976, and September 30, 1977*

Table 17-3

Numbers of Scaled Science Long-Term Trend Items Common Across Ages

Age	Booklets	Number of Items
9 only	91-93	55
13 only	91-93	30
17 only	84-85	32
9 and 13 only	91-93, 91-93	0
9 and 17 only	91-93, 84-85	0
13 and 17 only	91-93, 84-85	45*
9, 13, and 17	91-93, 91-93, 84-85	1
Total		**163**

******Note*: *One of these items (N406303) was treated as a different item from 1990 in the scaling of the 1992 assessment, but only for age 13. It was treated as an item common to 1992 and 1994 for all ages in the 1994 assessment.*

Table 17-4

Numbers of Scaled Science Long-Term Trend Items Common Across Assessments

	Number of Items		
Assessment Years	**Age 9**	**Age 13**	**Age 17**
1986, 1990, 1992, 1994	56	76	78
1982, 1986, 1990, 1992, 1994	10*	58	47
1977, 1986, 1990, 1992, 1994	56	76	76
1977, 1982, 1986, 1990, 1992, 1994	10*	58**	45

**Note: Twenty-four items common to years 1977 and 1982, but not later years, were included in the 1986 scaling of these items to stabilize the estimation of the item parameters. See* Expanding the New Design: The NAEP 1985-86 Technical Report *for more information.*

****Note: One of these items (N406303) was treated as a different item from 1990 in the scaling of the 1992 assessment, but only for age 13. It was treated as an item common to 1992 and 1994 for all ages in the 1994 assessment.*

17.2 ITEM ANALYSIS FOR THE SCIENCE LONG-TERM TREND ASSESSMENT

Conventional item analyses did not identify any difficulties with the 1994 long-term trend data for the 1994 samples that bridge to 1986. Table 17-5 contains information about the science long-term trend blocks. These blocks were presented to samples 9[MS-LTTrend], 13[MS-LTTrend], and 17[MS-LTTrend]. At ages 9 and 13, the blocks labeled S1, S2, and S3 were presented intact to 1986, 1990, 1992, and 1994 long-term trend samples. The age 9 and age 13 blocks appeared in booklets 91 through 93. At age 17, S1, S2, and S3 were presented intact to the 1986, 1990, 1992, and 1994 long-term trend samples. Block S3 was in booklet 84 and blocks S1 and S2 were in booklet 85. Table 17-6 shows the relationships between the blocks and booklets. Common labeling of these blocks across ages does not denote common items.

Table 17-5 contains the number of items, size of the sample administered the block, mean weighted proportion correct, mean weighted r-biserial, and mean weighted alpha as a measure of reliability for each block. The proportion of students attempting the last item in each block was very close to 1, due to audiotape administration. The average values were calculated using examinee sampling weights and the items in the block that were scaled. On average, the 1994 item-level statistics were not very different from those for the 1992 assessments. There were a number of individual items that had lower weighted proportion-correct values for the 1994 sample than for the 1992 sample of students. One of these, item N406303, was administered to age 13 students in block S1. This item is discussed in Section 17.3. Other items had higher weighted proportion-correct values for the 1994 sample, and there was a noticeable, if not significant tendency for average weighted r-biserials to be slightly higher in 1994 than in 1992. These variations did not significantly affect the estimation of item parameters using data from the 1992 and 1994 assessments. The percent of examinees not reaching items in the long-term trend blocks was always zero because the items were administered with a tape-recording to pace response time.

Table 17-5

Descriptive Statistics for Item Blocks in the Science Long-Term Trend Samples (1994)

AGE 9			
	Blocks		
Statistic	**S1**	**S2**	**S3**
Number of scaled items	17	20	19
Number of scaled constructed-response items	0	0	0
Unweighted sample size	1,892	1,878	1,893
Average weighted proportion correct	0.64	0.58	0.70
Average weighted r-biserial	0.58	0.49	0.58
Weighted alpha reliability	0.70	0.65	0.71

AGE 13			
	Blocks		
Statistic	**S1**	**S2**	**S3**
Number of scaled items	23	30	23
Number of scaled constructed-response items	0	0	0
Unweighted sample size	2,040	1,953	2,059
Average weighted proportion correct	0.54	0.56	0.61
Average weighted r-biserial	0.53	0.49	0.51
Weighted alpha reliability	0.74	0.77	0.70

AGE 17			
	Blocks		
Statistic	**S1**	**S2**	**S3***
Number of scaled items	24	31	23
Number of scaled constructed-response items	0	0	0
Unweighted sample size	1,924	1,924	1,889
Average weighted proportion correct	0.66	0.65	0.59
Average weighted r-biserial	0.50	0.54	0.62
Weighted alpha reliability	0.70	0.77	0.81

BEST COPY AVAILABLE

The correspondence between blocks, booklets, and samples is given for the long-term trend assessment in Tables 4-17 through 4-19 in Chapter 4. Common labeling of these blocks across ages does not denote common items.

17.3 IRT SCALING FOR THE SCIENCE LONG-TERM TREND ASSESSMENT

17.3.1 Item Parameter Estimation

The first step in the scaling process (described in Chapter 11) was the estimation of item parameters for the long-term trend items. This item calibration was performed using the NAEP version (Rogers & Nelson, 1990) of the BILOG/PARSCALE programs (Mislevy & Bock, 1982; Muraki & Bock, 1991) separately for each of the three age groups, using combined data from the 1990 and 1992 assessment years and treating each assessment sample as a sample from a separate subpopulation. The calibration was performed on all examinees using student weights during the entire scaling process. The weights for the 1992 samples were used in a rescaled form, where the sum of the rescaled weights for the 1992 samples was equal to the sum of the weights for the 1994 samples.

Item parameters were estimated separately for items at each age using 1992 and 1994 data with the NAEP-BILOG/PARSCALE computer program. Although other items were examined for irregularities, only items that were deleted from the previous scaling of the paced-tape long-term trend data were excluded in the 1992 analysis. Eleven percent of the items (18 items) administered to the long-term trend sample were excluded from analyses of previous assessments. The deleted items appear in Table 17-6. As a result of these deletions, 56 items were scaled for age 9, 76 items were scaled for age 13, and 78 items were scaled for age 17. A list of the items scaled for each of the ages, along with their item parameter estimates, appears in Appendix D.

Table 17-6

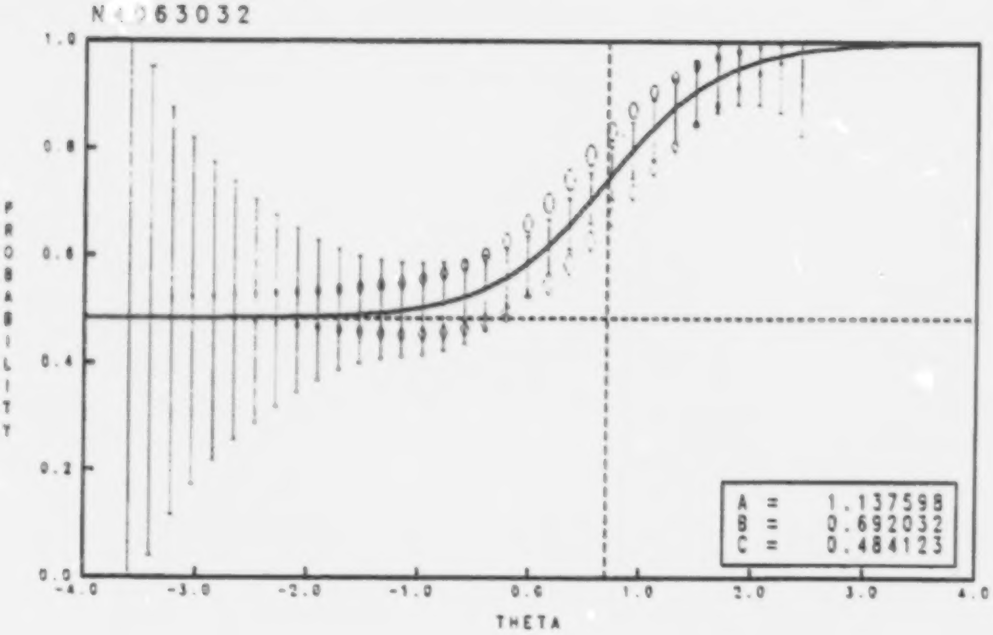
Items Deleted from the Paced-Tape Science Long-Term Trend Analysis

Age	Booklet	Block	Item	Reason for Exclusion
9	91	S1	N400201	Excluded in previous assessment
	92	S2	N401701	Excluded in previous assessment
	92	S2	N402003	Excluded in previous assessment
	92	S2	N402004	Excluded in previous assessment
	92	S2	N402601	Excluded in previous assessment
	92	S2	N402603	Excluded in previous assessment
	93	S3	N403802	Excluded in previous assessment
13	91	S1	N404902	Excluded in previous assessment
	91	S1	N404903	Excluded in previous assessment
	92	S2	N407501	Excluded in previous assessment
	93	S3	N409401	Excluded in previous assessment
	93	S3	N409402	Excluded in previous assessment
	93	S3	N409403	Excluded in previous assessment
	93	S3	N409801	Excluded in previous assessment
17	85	S1	N410001	Excluded in previous assessment
	85	S1	N410002	Excluded in previous assessment
	85	S1	N410301	Excluded in previous assessment
	85	S2	N407402	Excluded in previous assessment

The item response function, as well as the weighted proportion correct (as mentioned in Section 17.2), for one item (N4063032) presented to age 13 students in block S2 was different across the two assessment years included in the scaling. Figure 17-1 contains the empirical and model-based item response functions for this item, as estimated in 1992 from 1990 and 1992 data. The solid curve represents the model-based item response function, while the diamonds represent the empirical item response function for 1990 and the ovals represent the item response function for 1992. A description of functions of this type are given in Chapter 9. Because the empirical item response functions for the two assessment years were so different, this item was treated as if it were a different item in each of the two years. It was not considered to be an item that was common to both the 1990 and 1992 assessments, as all of the other items were. Figure 17-2 demonstrates that in 1994, the same item (concerning recycling), which had become strikingly easier from 1990 to 1992 (the 1992 ovals in Figure 17-1 appearing well above the 1990 diamonds), had moved back to a difficulty level slightly below the 1992 curve, but almost coincident with it, and no longer had to be regarded as showing changes inconsistent with those exhibited by the overall scale. Thus it was considered to be a common item between 1992 and 1994.

Figure 17-1

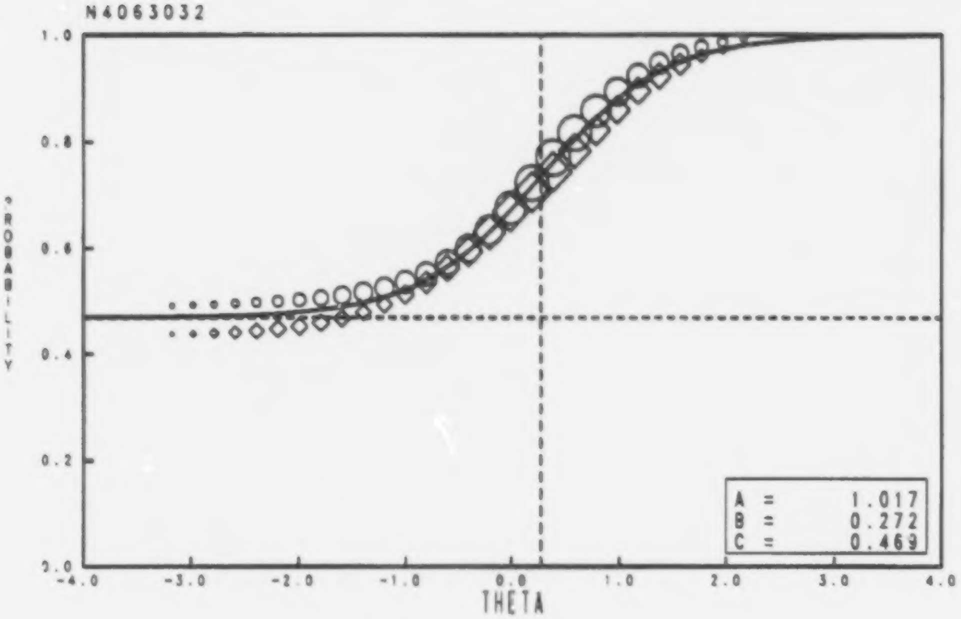
Example Long-Term Trend Item (N4063032, Age 13)
*Demonstrating Differential Item Functioning Across Assessment Years 1990 and 1992**



***Note:** The plot compares empirical and model-based estimates of the item response function (IRF). The smooth curve represents the model-based estimate at each provisional proficiency level. The diamonds represent 1992 data; ovals represent 1994 data.*

Figure 17-2

Example Long-Term Trend Item (N4063032, Age 13)
*Exhibiting Good Model Fit Across Assessment Years 1992 and 1994**



****Note:*** *The plot compares empirical and model-based estimates of the item response function (IRF). The smooth curve represents the model-based estimate at each provisional proficiency level. The diamonds represent 1992 data; ovals represent 1994 data.*

17.3.2 Derived Background Variables

In the long-term trend analysis, all variables derived for the science analysis from background questions were used both in generating plausible values and in reporting (to define subgroups). Derived conditioning and reporting variables are described in Appendix B.

17.4 GENERATION OF PLAUSIBLE VALUES

The generation of plausible values was conducted independently by age for each of the three assessment years. The item parameters from NAEP-BILOG/PARSCALE, final student weights, item responses and selected background variables (conditioning variables) were used with the computer program BGROUP (described in Chapter 11) in order to generate the values for each age. There were 45 contrasts in the conditioning model at age 9, including an overall constant, 48 at age 13, and 54 at age 17. Appendix C gives the codings for the conditioning variables for the three age groups. A check on the distributions of the plausible values for each age was made. The generation of plausible values is described in more detail in Chapters 9 and 11. Table 17-7 shows the proportion of variance accounted for by the conditioning model for each age.

Table 17-7
Proportion of Proficiency Variance Accounted for by the Conditioning Model for the Science Long-Term Trend Assessment

Age	Number of Conditioning Contrasts*	Proportion of Proficiency Variance
9	51	0.330
13	54	0.411
17	60	0.547

**__Note__: Excluding the constant term.*

17.5 THE FINAL SCIENCE LONG-TERM TREND SCALE

The linear indeterminacy of the long-term trend scale was resolved by linking the 1994 long-term trend scales to the previous long-term trend scales using the following procedure. For each age, the item parameters from 1994 based on data from 1992 and 1994 were used with the 1992 data to find plausible values for the 1992 data. The mean and standard deviation of all of the plausible values were calculated and matched to the mean and standard deviation of all of the plausible values based on the 1992 item parameters and 1992 data as reported in earlier reports. The transformations that resulted from this matching of the first two moments for the 1992 data are

$$\text{Age 9:}\quad \theta_{\text{proficiency}} = 37.38 \bullet \theta_{\text{calibrated}} + 232.95,$$

$$\text{Age 13:}\quad \theta_{\text{proficiency}} = 39.30\,e \bullet \theta_{\text{calibrated}} + 256.51, \text{ and}$$

$$\text{Age 17:}\quad \theta_{\text{proficiency}} = 48.20 \bullet \theta_{\text{calibrated}} + 292.99$$

where $\theta_{\text{proficiency}}$ denotes values on the final transformed scale and $\theta_{\text{calibrated}}$ denotes values on the original calibration scale. Overall summary statistics for the long-term trend samples are given in Table 17-8.

Table 17-8
Means and Standard Deviations on the
Science Long-Term Trend Scale

		All Five Plausible Values	
Age	**Assessment**	**Mean**	**S. D.**
9	1977	219.9	44.9
	1982	220.8	40.9
	1986	224.3	41.6
	1990	228.7	40.2
	1992	230.6	39.9
	1994	231.0	40.9
13	1977	247.4	43.5
	1982	250.1	38.6
	1986	251.4	36.6
	1990	255.2	37.6
	1992	258.0	36.9
	1994	256.8	37.2
17	1977	289.5	45.0
	1982	283.3	46.7
	1986	288.5	44.4
	1990	290.4	46.2
	1992	294.1	44.7
	1994	294.0	45.6

The main NAEP science composite scale was anchored in 1986, using the process described in *Expanding the New Design: The 1985-86 Technical Report* (Beaton, 1988). Because each of the 1994 scales was tied to the 1986 main cross-sectional or long-term trend scale through the 1990 and 1992 data, the distribution of proficiency scores derived from the main and long-term trend samples can be described in terms of scale anchors. In 1986 the levels of science proficiency were

150 = Knows everyday science facts;
200 = Understands simple scientific principles;
250 = Applies basic scientific information;
300 = Analyzes scientific procedures and data; and
350 = Integrates specialized scientific information.

17.6 PARTITIONING OF THE ESTIMATION ERROR VARIANCE

The variance of proficiency means for each grade was partitioned into the part due to the sampling of students and the part due to the latency of proficiency, θ, as described in Chapter 11. These estimates are given in Table 17-9 (for stability of the estimates, they are based on 100 plausible values). More detailed information for gender and race/ethnicity subgroups is available in Appendix E.

Table 17-9
Estimation Error Variance and Related Coefficients
for the Science Long-Term Trend Assessment

Age	Total Estimation Error Variance	Proportion of Variance Due to ...	
		Student Sampling	**Latency of θ**
9	1.01	0.84	0.16
13	1.08	0.90	0.10
17	2.39	0.94	0.06

Chapter 18

DATA ANALYSIS FOR THE LONG-TERM TREND WRITING ASSESSMENT[1]

James E. Carlson, Bruce A. Kaplan, Eiji Muraki, and Steve Wang
Educational Testing Service

18.1 INTRODUCTION

This chapter describes analyses of the writing prompts[2] and background items in the 1994 long-term trend assessment of writing. These analyses led to the results reported in the *NAEP 1994 Trends in Academic Progress: Achievement of U.S. Students in Science, 1969 to 1994; Mathematics, 1973 to 1994; Reading, 1971 to 1994; and Writing, 1984 to 1994* (Campbell, Reese, O'Sullivan, Dossey, & Donahue, 1996). Emphasis is given to the psychometric methods used to develop the composite scores that formed the basis of those reports.

The objectives of the writing analyses were to prepare scale values and perform all analyses necessary to produce a long-term trend report in writing. The writing long-term trend results include the years 1984, 1988, 1990, 1992, and 1994. Table 18-1 describes the samples of these years.

The 1994 samples used for the analysis of writing achievement are presented in Table 18-2. These samples comprise students selected both on the basis of age and grade in school. For the writing long-term trend analysis, unlike other subject-area long-term trend analyses, only those students selected on the basis of grade were included; the analysis sample is referred to as a "grade-only" sample.

As in the 1992 writing assessment, the IRT scaling for the 1994 assessment used a model for polytomously scored items. The 1992 assessment established the first such scale. The 1994 assessment represented the first time that the long-term trend writing scale had to be linked to a previously established scale. Sampling weights, as described in Chapter 10, were used for all analyses of the writing prompts.

---

[1] Bruce A. Kaplan was the primary person responsible for the planning, specification, and coordination of the writing long-term trend analyses. James E. Carlson and Eiji Muraki advised and reviewed the scaling of the long-term trend writing data. Computer activities for all long-term trend writing scaling and data analyses were directed and performed by Steve Wang with assistance from Norma A. Norris. Statistical advice was given by Eugene Johnson.

[2] The terms "item" and "prompt" are used interchangeably in this chapter.

Table 18-1

NAEP Writing Samples Contributing to 1994 Long-Term Trend Results, 1984-1994

Cohort	Year	Sample	Subjects	Time of Testing	Mode	Age Definition	Modal Grade
Age 9/ Grade 4	1984	Main	RW	Winter, spring	Print	CY	4
	1988	LTTrend*	RW	Winter	Print	CY	4
	1990	LTTrend*	RW	Winter	Print	CY	4
	1992	LTTrend*	RW	Winter	Print	CY	4
	1994	LTTrend*	RW	Winter	Print	CY	4
Age 13/ Grade 8	1984	Main	RW	Winter, spring	Print	CY	8
	1988	LTTrend*	RW	Winter	Print	CY	8
	1990	LTTrend*	RW	Winter	Print	CY	8
	1992	LTTrend*	RW	Winter	Print	CY	8
	1994	LTTrend*	RW	Winter	Print	CY	8
Age 17/ Grade 11	1984	Main	RW	Winter, spring	Print	Not CY	11
	1988	LTTrend*	RW	Winter	Print	Not CY	11
	1990	LTTrend*	RW	Winter	Print	Not CY	11
	1992	LTTrend*	RW	Winter	Print	Not CY	11
	1994	LTTrend*	RW	Winter	Print	Not CY	11

***Note**: *Within an age class group, these samples received common booklets and constituted a trend line.*

LEGEND

RW	*Reading and writing*
Main	*Main assessment*
LTTrend	*Long-term trend assessment*
Print	*Printed administration*
CY	*Calendar year: birthdates in 1984, 1980, and 1976 for ages 9, 13, and 17*
Not CY	*Age 17 only: birthdates between October 1, 1976, and September 30, 1977*

Table 18-2
NAEP Writing Student Samples

Sample	Booklets	Mode	Cohort Assessed	Time of Testing	Age Definition	Modal Grade	Number Assessed
9[RW-LTTrend]	51-56	Print	Age 9/ Grade 4	Winter	CY	4	5,335
13[RW-LTTrend]	51-56	Print	Age 13/ Grade 8	Fall	CY	8	5,547
17[RW-LTTrend]	51-56	Print	Age 17/ Grade 11	Spring	Not CY	11	4,840

LEGEND

RW	*Reading and writing*
LTTrend	*Long-term trend assessment*
Print	*Printed administration*
CY	*Calendar year: birthdates in 1984, 1980, and 1976 for ages 9, 13, and 17*
Not CY	*Age 17 only: birthdates between October 1, 1976, and September 30, 1977*

18.2 LONG-TERM TREND WRITING DATA ANALYSIS

When IRT scaling was used for the long-term trend writing assessment for the first time in 1992, a new scale was developed, and data from the 1984, 1988, and 1990 long-term trend samples were reanalyzed to establish the scale. The 1994 writing long-term trend points are based on data from the 1994 reading/writing long-term trend samples (RW-LTTrend in Table 18-2) that were linked to that scale. Each subsequent assessment matched the 1984 assessment in terms of the time of administration and age definitions. The booklets used in this assessment contained blocks of reading and writing items, as well as background questions. Identical booklets were used in 1984, 1988, 1990, 1992, and 1994.

The items on which the long-term trends in writing achievement are based are shown in Table 18-3. The table shows the block that contained the item in 1984 and long-term trend booklets containing the item in 1988, 1990, 1992, and 1994. Twelve writing prompts were used to measure long-term trends, with six prompts presented at each grade level. To allow comparisons in writing ability across grades, three of the six prompts presented to fourth-grade students were also presented to eighth-grade students; three of the eighth-grade prompts were also presented to eleventh-grade students; and one of the common prompts (Appleby House) was presented at all three grade levels.

Table 18-3

Assignment of 1984-94 Writing Long-Term Trend Items in 1984, 1988, 1990, 1992, and 1994

	1984 BIB-Spiral Blocks Used for Long-Term Trend Analyses			1988-94 Long-Term Trend Booklets		
Writing Task	**Grade 4**	**Grade 8**	**Grade 11**	**Grade 4**	**Grade 8**	**Grade 11**
N0003 Recreation Opportunity		C	C		52,54	52,54
N0004 Food on Frontier		D	D		51,54	51,54
N0005 Dissecting Frogs		E			53,55	
N0006 XYZ Company	E	E		52, 54	53,55	
N0009 Radio Station	G	G		54,55	55,56	
N0010 Appleby House	G	G	G	54,55	55,56	55,56
N0076 Flashlight	V*			56		
N0147 Plants	C			51,53		
N0148 Spaceship	E			52,54		
N0180 Space Program			E			53,55
N0190 Job Application			E			53,55
N0210 Bike Lane			G			55,56

**Note: Block V was not placed in a booklet with any other writing block in 1984 (all other blocks appeared with every other block at the same grade level in 1984), and hence could not be used in scaling.*

18.2.1 Primary Trait Scoring of the Writing Tasks and Measures of Scorer Reliability

All writing exercises from the 1994 assessment were scored for task accomplishment (primary trait). For the purposes of analysis, the student responses were coded as 0 (not rated), 1 (unsatisfactory), 2 (minimal), 3 (adequate), and 4 (elaborated). Not-reached and omitted items were excluded from the scaling. The writing long-term trend blocks contained either one or two items. If an item was left blank in a one-item block, it was scored as an omission. Items considered not-reached occurred only in writing blocks that had two cognitive items where the first item was answered and the second was not.

A 25 percent random subsample of all 1994 papers was rescored by a second rater to provide an estimate of interrater reliability. Although the measures of scorer agreement in NAEP have been consistently high, the possibility existed that there might be variation between the ratings provided by the group of scorers assembled in 1994 and the scorers assembled in previous years. Such a variation would be a confounding effect in trend measurement. The most direct way of controlling the effect of across-year variation in scoring would be to eliminate it entirely by rescoring all of the data from the previous four assessments, using the same set of scorers who scored the 1994 data. Unfortunately, resources did not allow for the rescoring of the full set of writing papers, but did allow for a rescoring of over 11,000 of the papers from 1988 (the numbers by prompt and grade are displayed in Table 18-4). The rescored papers for a given item constituted approximately 25 percent of all 1988 papers and consisted of all grade-eligible respondents to selected 1988 booklets containing that item. The procedure of rescoring data from other writing assessments was also used in the previous years' assessments in the current long-term trend analysis (1984, 1988, 1990, and 1992).

Because of rigorous training of scorers, it was expected that the between-year variability in scoring would be low enough to permit the use of the full set of the 1988 data. Table 18-4 shows scorer reliability, as measured by the intraclass correlation, for each prompt in the 1988, 1990, 1992, and 1994 data. The percentage of exact agreement between first and second raters is also given. In addition, the table shows the intraclass correlation and percentage of exact score agreement comparing the scores of samples of the 1990, 1992, and 1994 raters with those of the 1988 raters on a sample of the 1988 papers. The reliabilities and percentages of exact agreement (between first and second raters) were generally high for 1994 data, as they were for prior assessments (1984, 1988, 1990, and 1992).

18.3 ITEM ANALYSIS FOR THE WRITING LONG-TERM TREND ASSESSMENT

A standard item analysis for polytomous items, as described in Chapter 9, was conducted on the writing long-term trend item data. Table 18-5 displays the item analysis statistics for each grade—the number of examinees responding to each prompt, the percentage of examinees receiving each of the assigned scores, and the mean score of the prompt. R-polyserials and alpha reliabilities were not calculated since there is only one or two items per block.

The results of the item analysis were examined to verify that statistics for each item were in expected ranges. No difficulties were found in this process. Comparisons of item statistics with those of previous years were also made, and it was found that the items had similar statistics for all five years of the long-term trend in writing.

18.4 IRT SCALING FOR THE WRITING LONG-TERM TREND ASSESSMENT

This section describes the scaling of the primary trait data from the 1994 writing long-term trend assessment. A listing of the prompts used in scaling at the three grade levels is presented in Table 18-6. Five prompts were used at the fourth-grade level and six prompts were used at each of the eighth- and eleventh- grade levels. Either three or four scoring categories were used in the scaling for each of the prompts. Two prompts at the fourth-grade level and one at the eighth-grade level were scaled with three categories because the frequencies of responses in the fourth category were zero or near zero. All other prompts were scaled with four categories. After examining the pattern of omitted, not-reached, off-task, and illegible responses relative to responses to other prompts, it was decided to treat these responses as missing, because there appeared to be no obvious relationship between writing performance and nonresponse (for whatever reason) to other prompts. Treating such prompts as missing meant that they would not enter into the scaling process. One of the prompts administered in the assessment, "Flashlight," was not administered with any other prompt, and therefore could not be put on the same scale as the other prompts. Hence, this prompt was excluded from scaling and from the number of prompts cited above.

Table 18-4

Percentages of Exact Score Agreement and Interrater Reliability for the Primary Trait Scoring of the Writing Long-Term Trend Assessment Items

NAEP Item	1988 Data (by 1988 Raters)		1988 Data (by 1990 Raters)		1990 Data (by 1990 Raters)		1988 Data (by 1992 Raters)		1992 Data (by 1992 Raters)	
	Percent Agreement	Reliability	Percent Agreement	Reliability	Percent Agreement	Reliability	Percent Agreement	Reliability	Percent Agreement	Reliability
Age 9/Grade 4										
N0006 XYZ Company	97.1	.99	91.1	.90	88.8	.83	90.9	.92	88.0	.89
N0009 Radio Station	93.5	.95	89.0	.90	92.1	.93	92.0	.93	90.9	.93
N0010 Appleby House	90.3	.92	76.9	.78	78.5	.72	79.7	.83	78.4	.80
N0076 Flashlight	87.5	.88	80.5	.74	78.2	.77	76.6	.71	79.5	.75
N0147 Plants	94.3	.95	88.5	.89	82.4	.86	91.3	.92	88.5	.90
N0148 Spaceship	91.8	.95	83.7	.89	75.2	.82	77.8	.84	75.7	.79
Age 13/Grade 8										
N0003 Recreation Opportunity	85.4	.82	83.0	.81	76.7	.73	79.5	.77	81.9	.82
N0004 Food on Frontier	79.9	.68	83.5	.78	72.1	.67	79.4	.68	75.5	.69
N0005 Dissecting Frog	76.1	.64	80.6	.70	66.1	.56	71.2	.54	71.7	.63
N0006 XYZ Company	93.5	.92	92.6	.87	86.8	.76	86.2	.76	88.5	.80
N0009 Radio Station	87.0	.89	82.0	.79	80.7	.83	85.8	.87	86.0	.89
N0010 Appleby House	75.3	.69	75.4	.75	75.9	.72	78.0	.77	82.5	.84
Age 17/Grade 11										
N0003 Recreation Opportunity	90.8	.93	71.6	.78	76.3	.78	83.7	0.84	79.9	.78
N0004 Food on Frontier	93.1	.86	78.9	.69	76.7	.73	85.9	0.72	79.7	.73
N0010 Appleby House	89.3	.89	81.1	.81	81.6	.82	88.2	0.88	81.4	.82
N0180 Spaceship	89.9	.93	73.2	.75	71.8	.75	82.9	0.86	80.8	.81
N0190 Job Application	92.3	.92	85.5	.86	84.6	.83	90.2	0.88	90.5	.84
N0210 Bike Lane	84.9	.87	78.2	.76	75.6	.78	83.5	0.84	81.5	.83

(continued)

BEST COPY AVAILABLE

Table 18-4 (continued)

Percentages of Exact Score Agreement and Interrater Reliability
for the Primary Trait Scoring of the Writing Long-Term Trend Assessment Items

NAEP Item	1988 Data (by 1994 Raters)		1994 Data (by 1994 Raters)	
	Percent Agreement	Reliability	Percent Agreement	Reliability
Age 9/Grade 4				
N0006 XYZ Company	93.2	.92	94.4	.93
N0009 Radio Station	95.0	.96	94.9	.95
N0010 Appleby House	87.9	.89	90.4	.90
N0076 Flashlight	86.2	.82	81.0	.75
N0147 Plants	93.5	.94	89.4	.90
N0148 Spaceship	86.7	.90	84.0	.87
Age 13/Grade 8				
N0003 Recreation Opportunity	91.4	.88	90.1	.89
N0004 Food on Frontier	84.3	.79	87.6	.84
N0005 Dissecting Frog	84.4	.76	79.3	.71
N0006 XYZ Company	94.1	.92	95.6	.93
N0009 Radio Station	89.5	.90	95.5	.94
N0010 Appleby House	82.6	.80	83.4	.79
Age 17/Grade 11				
N0003 Recreation Opportunity	89.6	.92	82.0	.81
N0004 Food on Frontier	88.1	.74	82.7	.78
N0010 Appleby House	91.3	.89	87.4	.89
N0180 Spaceship	93.0	.92	77.0	.78
N0190 Job Application	93.2	.92	87.1	.84
N0210 Bike Lane	92.2	.92	82.2	.81

Table 18-5

Descriptive Statistics for Writing Prompts, 1994 Writing Long-Term Trend Samples

Item	RN	Weighted Percentages of Grade-Eligible Examinees in Each Score Category					Mean
		0*	1	2	3	4	
Grade 4							
N0006	1217	2.7	52.3	7.5	37.5	–	1.80
N0009	1303	3.4	50.0	31.3	15.8	0.0	1.60
N0010	1013	2.0	21.7	59.1	17.2	0.1	1.92
N0076	656	0.5	26.9	58.1	14.5	0.0	1.78
N0147	1347	1.1	15.4	47.5	36.1	–	2.19
N0148	1313	3.9	33.8	41.2	20.1	0.2	1.80
Grade 8							
N0003	1301	1.4	51.7	36.2	10.2	0.5	1.57
N0004	1321	1.9	22.0	63.9	12.0	0.2	1.87
N0005	1331	0.4	20.0	61.9	17.4	0.4	1.98
N0006	1313	0.2	20.1	6.7	73.1		2.53
N0009	1359	0.4	29.0	41.2	28.3	1.1	2.00
N0010	1255	0.2	6.7	49.7	42.8	0.7	2.37
Grade 11							
N0003	1272	1.4	34.7	40.4	21.6	2.0	1.88
N0004	1305	2.1	17.3	59.0	20.1	1.6	2.02
N0010	1141	1.2	10.3	25.4	59.6	3.5	2.54
N0180	1276	4.2	20.0	44.6	28.9	2.3	2.05
N0190	1214	0.5	16.4	19.3	60.5	3.3	2.50
N0210	1282	1.6	48.2	37.0	11.2	2.0	1.64

***Note**: *Omitted and off-task responses are coded as 0. For scaling, these responses are treated as not presented.*

Table 18-6

*Age/Grade Sample Sizes for Scaling Items Used in Long-Term Trends in Writing Performance**

NAEP Item	Grade 4					Grade 8					Grade 11				
	1984 Rescore	1988	1990	1992	1994	1984 Rescore	1988	1990	1992	1994	1984 Rescore	1988	1990	1992	1994
N0006 XYZ Company	603	1398	1679	2023	1488	751	1766	1992	1787	1776	–	–	–	–	–
N0009 Radio Station	582	1506	1794	2099	1577	713	1801	2025	1834	1831	–	–	–	–	–
N0010 Appleby House	530	1143	1413	1684	1217	674	1658	1848	1681	1681	694	1322	1628	1614	1344
N0147 Plants	709	1604	1833	2213	1708	–	–	–	–	–	–	–	–	–	–
N0148 Spaceship	664	1508	1765	2076	1587	–	–	–	–	–	–	–	–	–	–
N0003 Recreation Opportunity	–	–	–	–	–	598	1758	1962	1729	1739	635	1482	1767	1790	1542
N0004 Food on Frontier	–	–	–	–	–	731	1748	1951	1735	1749	714	1482	1756	1806	1547
N0005 Dissecting Frogs	–	–	–	–	–	777	1796	2016	1816	1803	–	–	–	–	–
N0180 Space Program	–	–	–	–	–	–	–	–	–	–	713	1464	1776	1777	1496
N0190 Job Application	–	–	–	–	–	–	–	–	–	–	728	1473	1796	1804	1472
N0210 Bike Lane	–	–	–	–	–	–	–	–	–	–	748	1495	1823	1801	1536

**Note*: *Omitted and off-task responses were treated as not presented.*

The resulting sample sizes for scaling the long-term trend items are provided in Table 18-6. The 1984 long-term trend point was based on a rescoring of a sample of the 1984 data by the 1988 raters. For details, see Johnson (1990).

When the long-term trend items were administered for the first time in 1984, they were used as part of the 1984 BIB design. By applying the additional information that could be obtained from the 1984 design, the long-term trend items for each grade were calibrated together. This task was not straightforward. Due to rater drift, the 1984 original data for long-term trend items had to be rescored in 1988 and not all the booklets were included in the rescoring. Although enough information was captured to calibrate all the fourth-grade items together, this was not the case for the eighth- and eleventh-grade items. There were two non-overlapping groups of items at these two grade levels. Items within the groups were administered to common samples of examinees, but there were no common samples for items between these two groups. This prohibited a direct linking via scaling all items simultaneously. As a result, an additional step was required for these two grade levels.

Fourth-Grade Level. For the fourth-grade level, the generalized partial credit model was used to calibrate, simultaneously, the prompts in the 1992 and 1994 data. This required a single run on the NAEP BILOG/PARSCALE computer program specifying two subpopulations, one for each year.

Eighth- and Eleventh-Grade Levels. For each of these two grade levels, two separate PARSCALE computer runs were conducted, one on each of the two separate groups of prompts discussed above. Like the fourth-grade level, the 1992 and 1994 data were simultaneously scaled using two subpopulations.

18.4.1 Item Parameter Estimation Using the Generalized Partial Credit Model

The first step in the scaling process was the estimation of item parameters for the long-term trend items. This item calibration was performed using the BILOG/PARSCALE program described in Chapter 11. Items were calibrated separately for each of the three age/grade groups. Item parameters were estimated using combined data from the assessment years 1992 and 1994, treating each assessment as a sample from a separate subpopulation. In age 13/grade 8 and age 17/grade 11, there were two sets of nonoverlapping items. Each set was scaled separately (see Table 18-7). Therefore, there were five sets of calibrations—one for age 9/grade 4, and two each for age 13/grade 8 and age 17/grade 11. Student weights were used for the calibration. To ensure that each assessment year had a similar influence on the calibration, student weights for 1992 and 1994 examinees were multiplied by a constant in order to adjust them to have the same sum, arbitrarily placed at 1,000. Approximately 1,200-1,800 examinees were present in each assessment year for each item.

Table 18-7

Nonoverlapping Sets of Items for the Writing Long-Term Trend Assessment

Age/Grade	Set	Booklets	Items
9/4	A	51-55	N0006, N0009, N0010, N0147, N0148
13/8	A	51, 52, 54	N0003, N0004
	B	53, 55, 56	N0005, N0006, N0009, N0010
17/11	A	51, 52, 54	N0003, N0004
	B	53, 55, 56	N0010, N0180, N0190, N0210

The final 1992 item parameter estimates were used as starting values. The proficiency distribution of each assessment year was constrained to be normally distributed, although the means and variances differed across assessment years. Calibration was concluded when changes in item parameters became negligibly small (i.e., less than .005). The item parameter estimates appear in Appendix D.

18.5 GENERATION OF PLAUSIBLE VALUES

The generation of plausible values was conducted independently for each set of items within an age/grade level for each of the assessment years. The item parameters from BILOG/PARSCALE, final student weights, item responses, and selected background variables were used with the computer program BGROUP (described in Chapter 11) to generate the values for each age. The background variables included student demographic characteristics (e.g., race/ethnicity of the student, highest level of education attained by parents), students' perceptions about writing, and student behavior both in and out of school (e.g., amount of television watched daily, amount of homework done each day). Appendix C gives the codings for the conditioning variables for the three age groups. Table 18-8 contains a list of the number of background variables included in conditioning, as well as the proportion of variance accounted for by the conditioning model for each age/grade.

Table 18-8

Proportion of Proficiency Variance Accounted for by the Conditioning Model for the Writing Long-Term Trend Assessment

Age/Grade	Number of Conditioning Constants*	Proportion of Variance
9/4	50	0.479
13/8	50	0.546
17/11	50	0.471

**Note: Excluding the constant term.*

18.6 THE FINAL WRITING LONG-TERM TREND SCALE

The linear indeterminacy of the long-term trend scale was resolved by linking the 1994 trend scales to previous trend scales. For sets within each age/grade, the item parameters from the joint calibration based on data from 1992 and 1994 were used with the 1992 data to find plausible values for the 1994 data. The mean and standard deviation for all of the plausible values were calculated and matched to the mean and standard deviation for all of the plausible values, based on the original analysis of the 1992 data, as given in earlier reports. The transformations that resulted from this matching of the first two moments for the 1992 data are

Age 9: $\theta_{\text{proficiency}} = 39.9466 \bullet \theta_{\text{calibrated}} + 200.1715$,

Age 13A: $\theta_{\text{proficiency}} = 37.9812 \bullet \theta_{\text{calibrated}} + 263.9819$,

Age 13B: $\theta_{\text{proficiency}} = 37.4556 \bullet \theta_{\text{calibrated}} + 265.8108$,

Age 17A: $\theta_{\text{proficiency}} = 35.4172 \bullet \theta_{\text{calibrated}} + 283.3398$, and

Age 17B: $\theta_{\text{proficiency}} = 32.6215 \bullet \theta_{\text{calibrated}} + 283.3663$,

where $\theta_{\text{proficiency}}$ denotes values on the final transformed scale and $\theta_{\text{calibrated}}$ denotes values on the calibration scale. Overall summary statistics for the long-term trend samples are given in Table 18-9.

Table 18-9
Means and Standard Deviations on the Writing Long-Term Trend Scale

Grade	Assessment Year	All Five Plausible Values	
		Mean	S.D.
4	1984	203.8	36.5
	1988	205.7	42.0
	1990	201.7	41.7
	1992	207.1	38.3
	1994	204.8	38.3
8	1984	266.7	29.5
	1988	263.7	32.4
	1990	256.6	37.5
	1992	274.4	36.3
	1994	265.3	35.9
11	1984	289.7	31.8
	1988	291.3	27.9
	1990	287.1	36.5
	1992	287.3	32.0
	1994	284.6	34.4

18.7 PARTITIONING OF ESTIMATION ERROR VARIANCE

The variance of proficiency means for each grade was partitioned into the part due to the sampling of students and the part due to the latency of proficiency, θ, as described in Chapter 11. These estimates are given in Table 18-10 (for stability of the estimates, they are based on 100 plausible values). More detailed information is available for gender and race/ethnicity subgroups in Appendix E.

Table 18-10
Estimation Error Variance and Related Coefficients for the Writing Long-Term Trend Assessment

Grade	Total Estimation Error Variance	Proportion of Variance Due to ...	
		Student Sampling	Latency of θ
4	1.61	0.48	0.52
8	1.08	0.59	0.41
11	1.39	0.65	0.35

18.8 ANALYSIS OF WRITING MECHANICS AND OVERALL FLUENCY

In addition to long-term trends in primary trait scores, trends were measured for the mechanics of writing and for overall writing fluency. Long-term trends in the mechanics of writing at each grade level were based on a selected writing prompt given to each grade level in 1984, 1988, 1990, 1992, and 1994. The items used for the assessment of the mechanics of writing were "Spaceship" (N0148) for grade 4 and "Recreation Opportunity" (N0003) for grades 8 and 11. All analyses were based on subsamples of approximately 500 responses to each item at each grade and year. Black students were sampled at a higher rate in order to provide sufficient sample sizes to allow for comparisons in performance between Black and White students. Student weights were adjusted to reflect the oversampling by a poststratification process. For each grade, the students selected for the writing mechanics analysis were categorized by gender and by race/ethnicity (White, Black, Hispanic, and Other), producing eight cells. The sampling weights of the students within each cell were then multiplied by a poststratification factor computed as a ratio whose denominator was the sum of weights of all students in the cell selected for the mechanics analysis and whose numerator was the sum of the weights of all students in the writing assessment of the specified grade, gender, and race/ethnicity. All papers used in this analysis were scored in 1994; the actual sample sizes are shown in Table 18-11.

Table 18-11
Sample Sizes for Mechanics Scoring

Grade	1984	1988	1990	1992	1994
4	506	484	567	678	521
8	474	517	601	563	530
11	522	497	602	566	509

Two writing items for each grade were scored holistically for overall writing fluency. To allow the measurement of long-term trends in overall writing fluency, most responses in the "grade-only" sample in the 1984, 1988, 1990, and 1992 assessments for the same items were also scored holistically in 1994. Table 18-12 shows the sample sizes for the measurement of long-term trends in the fluency of writing.

Table 18-12
Sample Sizes for Holistic Scoring

Grade	NAEP Item	Sample Size				
		1984	1988	1990	1992	1994
4	Flashlight	2021	611	702	840	634
	Spaceship	2026	1258	1356	1574	1204
8	Food on Frontier	2235	1341	1502	1308	1281
	Recreation Opportunity	2236	1372	1503	1309	1257
11	Food on Frontier	2372	1212	1399	1436	1212
	Recreation Opportunity	2349	1242	1415	1406	1144

The results of these analyses are reported in the *NAEP 1994 Trends in Academic Progress: Achievement of U.S. Students in Science, 1969 to 1994; Mathematics, 1973 to 1994; Reading, 1971 to 1994; and Writing, 1984 to 1994* (Campbell, Reese, O'Sullivan, Dossey, & Donahue, 1996).

Chapter 19

STATISTICAL SUMMARY OF THE 1994 NAEP SAMPLES[1]

Bruce A. Kaplan and Phillip T. Leung
Educational Testing Service

19.1 INTRODUCTION

The analysis of the 1994 NAEP data has resulted in the production of many thousands of tables presenting estimates of the proficiency of students, and various subgroups of students, in American schools. This chapter provides a statistical summary of the 1994 NAEP national samples. The chapter assumes a general familiarity with the structure of NAEP as summarized in the Introduction and in the overviews presented in Chapters 1 and 9. Similar results for the state samples appear in the compendium

Three of the many types of NAEP results are presented here:

- results of the instrument development process, including the sizes of the item pools and numbers of booklets;
- results of the sampling process, including the numbers of students in each sample by selected subgroups; and
- IEP and LEP counts and percentages in various samples.

19.2 MEASUREMENT INSTRUMENTS

For the 1994 assessment, 65 different assessment booklets and questionnaires were printed for age class 9, 87 for age class 13, and 85 for age class 17. These instruments are shown by age level and type in Table 19-1.

The item pool used to develop all main and long-term trend booklets is described in Table 19-2. In general, there are two types of items, cognitive and noncognitive. The cognitive items are developed to measure proficiency in particular subject areas, such as reading and mathematics. Cognitive items may be constructed-response or multiple-choice. The noncognitive items are usually questions about the student's or teacher's backgrounds and perceptions but may also probe other areas, such as school policies or teaching methods. Because many items were used at more than one age class, the total number of items in an item pool is not the sum of the item pools used for the three age classes.

---

[1] Bruce A. Kaplan was responsible for the text, specifying the tables, and coordinating table production. Phillip T. Leung produced most of the tables in this chapter. David Freund's advice was invaluable in the production of this chapter.

Table 19-3 shows the number of cognitive items in each subject area that were used in the separate samples.

The IEP/LEP student questionnaires, teacher questionnaires, and school characteristics and policies questionnaires contained only noncognitive questions. The number of items in the noncognitive pools is the same as the number of items in the questionnaires. More information about the instruments that were developed is provided in Chapters 2 and 4.

19.3 SAMPLE CHARACTERISTICS

In this section, the characteristics of the final NAEP samples are described. The process by which the samples were selected is discussed in Chapter 3.

In the 1994 main assessment, NAEP contacted 1,959 schools (1,952 original and seven replacements), of which 1,361 contributed data to the assessment. The disposition of these schools is shown in Table 19-4. Some of the schools were unwilling to cooperate; others were believed to be eligible from the sampling frame, but were not. The cooperation rate is calculated as the sum of cooperating schools and the schools that were found to have no eligible students divided by the same sum plus the schools that refused or were from districts that refused to cooperate.

Table 19-4 also shows the number of schools in several categories: region of the country (Northeast, Southeast, Central, West), school governance (public, nonpublic, Catholic, Bureau of Indian Affairs, Department of Defense), type of location, number of teachers, and number of students.

For the 1994 long-term trend studies, NAEP contacted 869 schools (864 original and five replacement), of which 678 contributed data to the various trend assessments. Table 19-5 supplies the same information for the schools assessed for the long-term trend studies that Table 19-4 supplies for the main assessment schools.

The numbers of respondents to the teacher questionnaires are summarized in Table 19-6. The first column in this table includes the number of teachers who responded, by grade and subject area. The second column is the number of students who were not linked to teachers. The third column is the number of students linked to teachers, but not specific classes of these teachers, for eighth grade, or teachers who did not answer classroom information for fourth grade. The last column is the number of students linked to their teachers and their specific classes.

NAEP is administered in units called assessment sessions. If the number of students attending an assessment session is fewer than a predetermined number, the students missing from the session are assigned to a makeup session and then assessed. Table 19-7 shows the number of regular and makeup sessions in 1994 NAEP by age class for the main and long-term trend samples. Altogether, 143,738 assessed and excluded students were involved in the 1994 NAEP. The breakdown by age class and by sample is shown in Table 19-8. Note that in 1994, rules for makeup sessions were tighted up from previous years, at NCES's request.

Tables 19-9 through 19-17 display the distribution of the students assessed in the main NAEP assessment in several basic categories: gender, racial/ethnic grouping, region of the country, parental education, and type of location. There is one table for each main sample taken: reading, U.S. history, and geography for each of the three age classes. These tables have four columns:

- eligible by age, which means that the students were in an appropriate age group;
- eligible by grade, which means that the students were in an appropriate grade;
- eligible by age and by grade, which means that the students were of both an appropriate age and appropriate grade; and
- eligible by age or by grade, which is the total number of students for whom data were collected.

Tables 19-18 through 19-23 contain the distribution of students in the same categories by age class for the long-term trend samples. Tables 19-18 to 19-20 contain the distributions for the reading and writing long-term trend samples. Tables 19-21 to 19-23 display the distributions for the mathematics and science long-term trend samples.

Similarly, Tables 19-24 through 19-26 contain the distribution of excluded students by age class for the main samples. Tables 19-27 through 19-29 enumerate the excluded students across the various long-term trend samples.

19.4 POPULATION ESTIMATES

The 1994 NAEP samples were designed for estimating the size and attributes of a number of different populations of students. The estimation procedures use sampling weights, developed by Westat, Inc., that are used in conjunction with the members of the sample (see Chapter 3). In this chapter, all estimates of population parameters use these sampling weights.

Table 19-31 shows the sizes of the various samples and the estimated population sizes by age/grade. The sum of the initial weights for a given sample is an estimate of the number of students who are in the population represented by the sample. In other words, the sum of the initial weights is taken as the estimated population size. In analyses, however, this sum of weights was rescaled to sum to the sample size.

Due to design considerations the main assessment was divided into subsamples, and were administered, and therefore weighted, independently, so that the sum of the initial weights for each subsample estimates the population size. The subsamples were reading, U.S. history, and geography for all three age classes.

Note that the samples for the main assessment and the samples for all three age classes of the reading and writing long-term trend are grade and age samples. The samples for the mathematics and science long-term trend are age-only samples. The sum of the initial weights of the excluded students estimates the number of ineligible students at the respective age/grade levels.

In most cases, the number of students in an age/grade combination is not of interest; a researcher will be interested in estimating the number of students at either a grade or an age level. For the samples that contain both grade- and age-eligible students, an estimate of the total number of students at an age level can be made by summing the initial weights of only the age-eligible students and adding the corresponding sample of age-eligible excluded students' initial weights. An estimate of the total number of students in a grade sample can be made by summing the initial weights of grade-eligible students plus the initial weights of grade-eligible students from the appropriate excluded student sample.

The next group of tables estimates how many students in the main NAEP samples are age-eligible and grade-eligible by age class. Tables 19-31 through 19-39 show how many students at a particular grade level are at, in, or above the modal age for that grade, and how many at a particular age level are at, in, or above the modal grade for that age. Along with the counts from these samples are estimates of the numbers of students in these categories in the population. The standard errors of these estimates and coefficients of variation are also given. (See Chapter 9 for an explanation of coefficient of variation.)

Tables 19-40 through 19-43 contain similar information for the long-term trend samples, by age level. Where age-only samples are shown, information for all three ages is given in one table, since the partitioning of the sample by modal age groupings provides no added information.

Tables 19-44 to 19-64 show the sizes of the estimated populations of assessable students and the weighted percentages for the NAEP reporting categories of gender, race/ethnicity, region of the country, parents' education level, and type of location. The estimated subpopulation percentages for the main NAEP samples are shown in Tables 19-44 through 19-52, separately by age eligibility, grade eligibility, and age/grade eligibility. Tables 19-53 to 19-58 show the same information for the long-term trend samples. In a similar manner, Tables 19-59 to 19-64 show the estimated total population of excluded students and the weighted percentages by demographic subgroups (data about parents' education level is not collected for excluded students and therefore not reported; data about reasons for exclusion are included instead).

Students were assigned proficiency values in a subject area only if they received at least one assessment block in that subject that was scaled. Thus, the sample sizes of students who have proficiency values vary from one subject area to another. Tables 19-65 and 19-66 show the number of students with proficiency values in each subject area by age and grade combinations. Table 19-65 is for the main assessment; Table 19-66 is for the long-term trend assessments.

Tables 19-67 to 19-72 deal with the IEP (Individual Education Program) and LEP (Limited English Proficiency) students. The first three tables, one for each grade/age, have raw counts of the number of IEP and LEP students for various samples—assessed students, excluded students, assessed and excluded students. There are also raw counts for the number of excluded and assessed students for IEP students, LEP students, either IEP or LEP students, and both IEP and LEP students. Note that the "Assessed Students" category represents the number assessed across all three main assessment subject areas. Tables 19-70 to 19-72, one for each grade/age, are weighted percentages and total population estimates for the same samples as in Tables 19-67 to Tables 19-69. To correctly calculate the estimates of percents and totals, the assessed students were used with a weight of one-third their original weight (ORIGWT). This was done because the three samples of assessed students were combined.

In previous years, this chapter also provided several tables showing selected proficiency results for assessed students, as an aid to readers who are interested in the estimates of proficiency that led to the interpretive results provided in the NAEP subject area reports. These tables are no longer included in this report. Instead, readers are encouraged to take advantage of the electronic version of these results, including thousands of tables computed to analyze the 1994 data, provided through the NAEP Almanac CD-ROM and Viewer navigational software. The Almanac Viewer is part of the NAEP Data on Disk product suite, which also includes a complete set of secondary-use data files, including state samples, on CD-ROM, PC-based NAEP data extraction software, and NAEP analysis modules. These products provide versatile access to a wealth of NAEP data.

Table 19-1

Measurement Instruments Developed for 1994 NAEP

	Age Class		
	9	13	17
Student Assessment Booklets			
TOTAL NUMBER OF CROSS-SECTIONAL (MAIN)	52	73	74
Reading	16	21	22
U.S. History	18	33	33
Geography	18	19	19
TOTAL NUMBER OF LONG-TERM TREND	9	9	8
Reading and Writing	6	6	6
Mathematics and Science	3	3	2
TOTAL NUMBER OF QUESTIONNAIRES	4	5	3
Excluded Student	1	1	1
Teacher	1	2	0
School	1	1	1
IEP/LEP	1	1	1
TOTAL NUMBER OF STUDENT ASSESSMENT BOOKLETS	**65**	**87**	**85**

Table 19-2

Number of Items Administered, by Sample and Age Class

	Age Class			Total Number of Distinct Items
	9	13	17	
COMMON BACKGROUND				
Cross-Sectional (Main)	22	21	30	40
Reading and Writing Long-Term Trend	37	37	48	48
Math and Science Long-Term Trend	28	30	48	58
READING MAIN				
Background	15	25	26	27
Cognitive	84	121	131	272
Motivation	5	5	5	5
U.S. HISTORY MAIN				
Background	17	23	25	29
Cognitive	94	148	156	331
Motivation	5	5	5	5
GEOGRAPHY MAIN				
Background	22	25	29	34
Cognitive	90	125	123	273
Motivation	5	5	5	5
READING LONG-TERM TREND				
Background	40	42	78	81
Cognitive	105	108	96	193
WRITING LONG-TERM TREND				
Background	53	65	65	77
Cognitive	6	6	6	12
MATHEMATICS LONG-TERM TREND				
Background	3	29	39	49
Cognitive	68	98	94	184
SCIENCE LONG-TERM TREND				
Background	16	29	29	45
Cognitive	63	83	82	180
IEP/LEP STUDENT QUESTIONNAIRE	66	66	66	66
EXCLUDED STUDENT QUESTIONNAIRE	67	67	67	67

Table 19-2 (continued)

Number of Items Administered, by Sample and Age Class

	Age Class			Total Number of Distinct Items
	9	13	17	
READING TEACHER QUESTIONNAIRE				
Teacher background	25	17	—	29
Reading background	15	13	—	15
Reading classroom	31	31	—	33
U.S. HISTORY TEACHER QUESTIONNAIRE				
Teacher background	25	17	—	30
U.S. History background	6	6	—	12
U.S. History classroom	35	35	—	33
GEOGRAPHY TEACHER QUESTIONNAIRE				
Teacher background	25	17	—	30
Geography background	9	8	—	14
Geography classroom	28	29	—	31
SCHOOL QUESTIONNAIRE	64	74	82	114

Table 19-3

Number of Distinct Cognitive Items, by Sample and Age Class

	Age Class			
	9	13	17	Total
CROSS-SECTIONAL (MAIN)				
Reading	84	124	134	288
U.S. History	94	148	156	331
Geography	90	125	123	273
LONG-TERM TREND				
Reading	105	108	96	193
Writing	6	6	6	12
Mathematics	68	98	94	184
Science	63	83	82	180
TOTAL* COGNITIVE ITEMS	**510**	**692**	**691**	**1461**

**Note*: *Because many items were used at more than one age class and/or for more than one sample, the total number of cognitive items is not equal to the total number of distinct items used for the three age classes and across the samples. Item counts in this table are for all items presented, but not necessarily used in scaling. Therefore, the number of items shown here may not agree with item counts reported in other chapters.*

Table 19-4
School Characteristics in Main Samples*

	Age/Grade 9/4	13/8	17/12	Total
TOTAL ORIGINAL SAMPLE	511	785	656	1952
Cooperating	396	537	421	1354
Out of range or closed	33	73	96	202
No eligibles enrolled	15	72	19	106
District refused	33	42	48	123
School refused	34	61	72	167
COOPERATION RATE	86	86	79	83
COOPERATING REPLACEMENTS FOR REFUSALS	3	2	2	7
TOTALS				
Cooperating schools	399	539	423	1361
Completed questionnaires	384	504	393	1281
REGION				
Northeast	94	124	89	307
Southeast	88	122	109	319
Central	110	136	97	343
West	107	157	128	392
SCHOOL TYPE				
Public	254	297	298	849
Nonpublic	68	114	84	266
Catholic	77	127	40	244
Bureau of Indian Affairs	0	1	1	2
Department of Defense	0	0	0	0

****Note***: *NAEP reporting subgroups are described in Appendix B.*

Table 19-4 (continued)

School Characteristics in Main Samples

	Age/Grade			
	9/4	13/8	17/12	Total
NUMBER OF TEACHERS				
Unclassified	0	0	0	0
1 - 4	7	15	4	26
5 - 9	59	64	13	136
10 - 19	112	135	53	300
20 - 49	174	194	130	498
50 - 74	18	61	68	147
75 - 99	1	13	58	72
100 +	2	3	59	64
Missing	15	35	30	80
NUMBER OF STUDENTS				
Unclassified	0	0	0	0
1 - 99	21	27	19	67
100 - 299	133	178	77	388
300 - 499	109	90	60	259
500 - 749	74	89	41	204
750 - 999	22	55	42	119
1000 - 1499	13	38	68	119
1500 +	1	8	78	87
Missing	15	35	30	80

Table 19-5

School Characteristics in Long-Term Trend Samples*

	Age Class			
	9	13	17	Total
TOTAL ORIGINAL SAMPLE	289	330	245	864
Cooperating	243	239	191	673
Out of range or closed	3	3	1	7
No eligibles enrolled	5	28	7	40
District refused	20	35	20	75
School refused	18	25	26	69
COOPERATION RATE	87	82	81	83
COOPERATING REPLACEMENTS FOR REFUSALS	0	1	4	5
TOTALS				
Cooperating schools	243	240	195	678
Completed questionnaires	236	234	176	646
REGION				
Northeast	45	48	44	137
Southeast	54	57	43	154
Central	68	56	51	175
West	76	79	57	212
SCHOOL TYPE				
Public	201	197	171	569
Nonpublic	20	15	11	46
Catholic	22	28	13	63
Bureau of Indian Affairs	0	0	0	0
Department of Defense	0	0	0	0

****Note***: *NAEP reporting subgroups are described in Appendix B.*

Table 19-5 (continued)

School Characteristics in Long-Term Trend Samples

	Age Class			
	9	13	17	Total
NUMBER OF TEACHERS				
Unclassified	0	0	0	0
1 - 4	2	2	0	4
5 - 9	22	14	3	39
10 - 19	50	41	18	109
20 - 49	152	124	44	320
50 - 74	9	41	44	94
75 - 99	1	6	29	36
100 +	0	5	38	43
Missing	7	6	19	32
NUMBER OF STUDENTS				
Unclassified	0	0	0	0
1 - 99	6	5	3	14
100 - 299	55	41	21	117
300 - 499	81	51	15	147
500 - 749	63	60	25	148
750 - 999	25	43	16	84
1000 - 1499	6	27	48	81
1500 +	0	6	48	54
Missing	7	6	19	32

Table 19-6

Numbers of Responses to Teacher Questionnaires and Students Matched with Teacher Data

	Number of Teachers Responding	Number of Students with		
		No Match	Partial Match	Complete Match
GRADE 4				
Reading	824	441	148	6793
U.S. History	826	312	333	4854
Geography	840	268	416	4823
GRADE 8				
Reading	1077	1835	1655	12116
U.S. History	786	1159	833	6775
Geography	716	1424	2528	2926

Table 19-7

Numbers of Assessment Sessions by Sample, Type of Session, and Age Class

	Age Class			
	9	13	17	Total
MAIN				
Regular sessions	1766	2517	2457	6740
Makeup sessions	20	67	232	319
READING AND WRITING LONG-TERM TREND				
Regular sessions	261	246	253	760
Makeup sessions	0	3	18	21
MATH AND SCIENCE LONG-TERM TREND				
Regular sessions	230	224	181	635
Makeup sessions	5	6	9	20
TOTAL				
Regular sessions	2257	2987	2891	8135
Makeup sessions	25	76	259	360

Table 19-8

Number of Students Assessed and Excluded
by Sample and Age Class

	Age Class			
	9	13	17	Total
ASSESSED STUDENTS				
Main	24090	41728	38946	104764
Reading and Writing Long-Term Trend	5335	5547	4840	15722
Math and Science Long-Term Trend	5663	6052	3813	15528
EXCLUDED STUDENTS				
Main	1512	2044	1874	5430
Long-Term Trend	984	806	504	2294
TOTAL	37584	56177	49977	143738

Table 19-9

Number of Students in the Reading Main Sample
by Type of Eligibility and Subgroup Classification, Age 9/Grade 4*

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
TOTAL	7042	7382	4744	9680
SEX				
Male	3580	3734	2268	5046
Female	3462	3648	2476	4634
RACE/ETHNICITY				
White	4293	4401	2836	5858
Black	1106	1261	814	1553
Hispanic	1136	1184	714	1606
Asian American	200	211	165	246
Pacific Islander	87	96	66	117
American Indian	161	170	102	229
Unclassified	59	59	47	71
REGION				
Northeast	1720	1816	1347	2189
Southeast	1807	1888	1130	2565
Central	1496	1571	907	2160
West	2019	2107	1360	2766
PARENTS' EDUCATION				
Less than high school	292	303	171	424
High school	806	889	538	1157
Greater than high school	471	540	347	664
Graduated college	2938	3134	2084	3988
Unknown	2518	2501	1598	3421
TYPE OF LOCATION				
Large city	977	1135	779	1333
Midsize city	1740	1804	1144	2400
Urban fringe of large city	1918	1892	1304	2506
Urban fringe of midsize city	1031	1039	671	1399
Large town	217	246	158	305
Small town	650	708	390	968
Rural/MSA	68	85	45	108
Rural/NonMSA	441	473	253	661

**Note*: *NAEP reporting subgroups are described in Appendix B.*

Table 19-10

Number of Students in the U.S. History Main Sample
by Type of Eligibility and Subgroup Classification, Age 9/Grade 4*

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
TOTAL	5239	5499	3533	7205
SEX				
Male	2565	2688	1627	3626
Female	2674	2811	1906	3579
RACE/ETHNICITY				
White	3179	3292	2089	4382
Black	910	1072	681	1301
Hispanic	778	761	488	1051
Asian American	147	142	115	174
Pacific Islander	71	75	53	93
American Indian	113	117	71	159
Unclassified	41	40	36	45
REGION				
Northeast	1311	1334	1005	1640
Southeast	1361	1440	853	1948
Central	1147	1241	731	1657
West	1420	1484	944	1960
PARENTS' EDUCATION				
Less than high school	195	196	104	287
High school	560	631	371	820
Greater than high school	353	402	264	491
Graduated college	2218	2428	1594	3052
Unknown	1891	1827	1189	2529
TYPE OF LOCATION				
Large city	864	919	675	1108
Midsize city	1310	1365	848	1827
Urban fringe of large city	1348	1393	950	1791
Urban fringe of midsize city	703	732	460	975
Large town	135	153	93	195
Small town	539	545	299	785
Rural/MSA	66	85	43	108
Rural/NonMSA	274	307	165	416

**Note: NAEP reporting subgroups are described in Appendix B.*

Table 19-11

Number of Students in the Geography Main Sample
by Type of Eligibility and Subgroup Classification, Age 9/Grade 4*

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
TOTAL	5167	5507	3469	7205
SEX				
Male	2559	2797	1636	3720
Female	2608	2710	1833	3485
RACE/ETHNICITY				
White	3163	3295	2059	4399
Black	880	1042	648	1274
Hispanic	767	805	499	1073
Asian American	124	138	102	160
Pacific Islander	76	76	58	94
American Indian	109	110	71	148
Unclassified	48	41	32	57
REGION				
Northeast	1287	1362	1015	1634
Southeast	1354	1445	841	1958
Central	1121	1216	682	1655
West	1405	1484	931	1958
PARENTS' EDUCATION				
Less than high school	191	211	116	286
High school	554	620	345	829
Greater than high school	327	388	241	474
Graduated college	2195	2386	1567	3014
Unknown	1874	1881	1190	2565
TYPE OF LOCATION				
Large city	823	928	666	1085
Midsize city	1288	1353	821	1820
Urban fringe of large city	1332	1387	910	1809
Urban fringe of midsize city	703	748	467	984
Large town	137	159	93	203
Small town	524	544	293	775
Rural/MSA	77	84	51	110
Rural/NonMSA	283	304	168	419

****Note:*** *NAEP reporting subgroups are described in Appendix B.*

BEST COPY AVAILABLE

Table 19-12

Number of Students in the Reading Main Sample
by Type of Eligibility and Subgroup Classification, Age 13/Grade 8*

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
TOTAL	14831	15606	9564	20873
SEX				
Male	7377	7736	4373	10740
Female	7454	7870	5191	10133
RACE/ETHNICITY				
White	9212	9640	6014	12838
Black	2361	2616	1496	3481
Hispanic	2215	2294	1324	3185
Asian American	424	438	329	533
Pacific Islander	267	280	207	340
American Indian	271	256	134	393
Unclassified	81	82	60	103
REGION				
Northeast	2912	2961	2132	3741
Southeast	4414	4794	2630	6578
Central	3036	3294	1953	4377
West	4469	4557	2849	6177
PARENTS' EDUCATION				
Less than high school	1032	1137	567	1602
High school	3047	3221	1861	4407
Greater than high school	2606	2915	1825	3696
Graduated college	6452	6748	4441	8759
Unknown	1624	1515	831	2308
TYPE OF LOCATION				
Large city	2330	2426	1625	3131
Midsize city	3595	3717	2259	5053
Urban fringe of large city	2761	2842	1905	3698
Urban fringe of midsize city	2620	2894	1797	3717
Large town	332	375	222	485
Small town	1612	1773	931	2454
Rural/MSA	512	532	278	766
Rural/NonMSA	1069	1047	547	1569

****Note:*** *NAEP reporting subgroups are described in Appendix B.*

Table 19-13

Number of Students in the U.S. History Main Sample by Type of Eligibility and Subgroup Classification, Age 13/Grade 8*

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
TOTAL	8302	8767	5400	11669
SEX				
Male	4037	4282	2406	5913
Female	4265	4485	2994	5756
RACE/ETHNICITY				
White	5093	5233	3322	7004
Black	1391	1574	895	2070
Hispanic	1213	1305	757	1761
Asian American	258	283	204	337
Pacific Islander	156	165	119	202
American Indian	154	159	71	242
Unclassified	37	48	32	53
REGION				
Northeast	1602	1614	1151	2065
Southeast	2502	2664	1511	3655
Central	1682	1821	1092	2411
West	2516	2668	1646	3538
PARENTS' EDUCATION				
Less than high school	569	656	310	915
High school	1743	1867	1061	2549
Greater than high school	1443	1608	1010	2041
Graduated college	3590	3721	2502	4809
Unknown	926	880	500	1306
TYPE OF LOCATION				
Large city	1331	1440	917	1854
Midsize city	2008	2106	1289	2825
Urban fringe of large city	1553	1568	1076	2045
Urban fringe of midsize city	1368	1494	949	1913
Large town	157	207	108	256
Small town	937	998	561	1374
Rural/MSA	310	315	168	457
Rural/NonMSA	638	639	332	945

****Note***: *NAEP reporting subgroups are described in Appendix B.*

Table 19-14

Number of Students in the Geography Main Sample
by Type of Eligibility and Subgroup Classification, Age 13/Grade 8*

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
TOTAL	6558	6878	4250	9186
SEX				
Male	3236	3410	1931	4715
Female	3322	3468	2319	4471
RACE/ETHNICITY				
White	4051	4140	2616	5575
Black	1147	1266	745	1668
Hispanic	911	990	568	1333
Asian American	183	187	140	230
Pacific Islander	112	117	87	142
American Indian	122	130	67	185
Unclassified	32	48	27	53
REGION				
Northeast	1306	1289	952	1643
Southeast	1964	2075	1160	2879
Central	1285	1444	835	1894
West	2003	2070	1303	2770
PARENTS' EDUCATION				
Less than high school	440	507	257	690
High school	1330	1438	793	1975
Greater than high school	1150	1276	829	1597
Graduated college	2857	2915	1953	3819
Unknown	759	717	405	1071
TYPE OF LOCATION				
Large city	1074	1144	748	1470
Midsize city	1621	1621	1017	2225
Urban fringe of large city	1224	1235	858	1601
Urban fringe of midsize city	1042	1174	717	1499
Large town	138	154	101	191
Small town	723	778	426	1075
Rural/MSA	247	270	138	379
Rural/NonMSA	489	502	245	746

****Note:*** *NAEP reporting subgroups are described in Appendix B.*

Table 19-15

Number of Students in the Reading Main Sample by Type of Eligibility and Subgroup Classification, Age 17/Grade 12*

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
TOTAL	14668	14181	9329	19520
SEX				
Male	7339	7144	4351	10132
Female	7329	7037	4978	9388
RACE/ETHNICITY				
White	9502	9327	6347	12482
Black	2330	2214	1317	3227
Hispanic	1930	1678	1032	2576
Asian American	483	538	345	676
Pacific Islander	242	233	182	293
American Indian	133	135	76	192
Unclassified	48	56	30	74
REGION				
Northeast	3266	3281	2411	4136
Southeast	4219	3962	2480	5701
Central	2965	2864	1852	3977
West	4218	4074	2586	5706
PARENTS' EDUCATION				
Less than high school	1222	1161	589	1794
High school	3004	2886	1752	4138
Greater than high school	3564	3461	2366	4659
Graduated college	6344	6199	4374	8169
Unknown	452	407	203	656
TYPE OF LOCATION				
Large city	2211	2082	1417	2876
Midsize city	3384	3246	2175	4455
Urban fringe of large city	3425	3419	2377	4467
Urban fringe of midsize city	1926	1878	1257	2547
Large town	317	321	210	428
Small town	1866	1736	1033	2569
Rural/MSA	535	512	294	753
Rural/NonMSA	1004	987	566	1425

**Note: NAEP reporting subgroups are described in Appendix B.*

Table 19-16

Number of Students in the U.S. History Main Sample
by Type of Eligibility and Subgroup Classification, Age 17/Grade 12*

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
TOTAL	8132	7818	5113	10837
SEX				
Male	4148	3933	2395	5686
Female	3984	3885	2718	5151
RACE/ETHNICITY				
White	5300	5155	3503	6952
Black	1321	1243	759	1805
Hispanic	1008	909	516	1401
Asian American	275	289	184	380
Pacific Islander	115	115	86	144
American Indian	77	75	44	108
Unclassified	36	32	21	47
REGION				
Northeast	1816	1770	1321	2265
Southeast	2307	2168	1346	3129
Central	1628	1575	977	2226
West	2381	2305	1469	3217
PARENTS' EDUCATION				
Less than high school	672	612	317	967
High school	1655	1502	914	2243
Greater than high school	1990	1959	1327	2622
Graduated college	3500	3473	2414	4559
Unknown	276	234	120	390
TYPE OF LOCATION				
Large city	1178	1135	763	1550
Midsize city	1755	1709	1099	2365
Urban fringe of large city	1934	1922	1325	2531
Urban fringe of midsize city	1118	1038	702	1454
Large town	192	179	124	247
Small town	1096	1014	617	1493
Rural/MSA	286	266	158	394
Rural/NonMSA	573	555	325	803

****Note***: *NAEP reporting subgroups are described in Appendix B.*

Table 19-17

Number of Students in the Geography Main Sample
by Type of Eligibility and Subgroup Classification, Age 17/Grade 12*

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
TOTAL	6473	6234	4118	8589
SEX				
Male	3272	3118	1920	4470
Female	3201	3116	2198	4119
RACE/ETHNICITY				
White	4195	4079	2780	5494
Black	1066	1003	622	1447
Hispanic	801	724	433	1092
Asian American	209	236	148	297
Pacific Islander	103	109	81	131
American Indian	72	60	40	92
Unclassified	27	23	14	36
REGION				
Northeast	1404	1407	1028	1783
Southeast	1839	1729	1083	2485
Central	1323	1275	820	1778
West	1907	1823	1187	2543
PARENTS' EDUCATION				
Less than high school	561	486	280	767
High school	1303	1307	777	1833
Greater than high school	1579	1536	1055	2060
Graduated college	2778	2692	1890	3580
Unknown	218	178	95	301
TYPE OF LOCATION				
Large city	942	916	630	1228
Midsize city	1456	1400	946	1910
Urban fringe of large city	1507	1477	1026	1958
Urban fringe of midsize city	885	822	556	1151
Large town	139	151	90	200
Small town	872	798	477	1193
Rural/MSA	239	223	131	331
Rural/NonMSA	433	447	262	618

**__Note__: NAEP reporting subgroups are described in Appendix B.*

Table 19-18

Number of Students in the Reading and Writing Long-Term Trend Sample by Type of Eligibility and Subgroup Classification, Age 9/Grade 4*

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
TOTAL	3871	4098	2634	5335
SEX				
Male	1936	2063	1259	2740
Female	1935	2035	1375	2595
RACE/ETHNICITY				
White	2370	2472	1588	3254
Black	500	597	371	726
Hispanic	786	787	510	1063
Asian American	99	115	83	131
American Indian	113	122	81	154
Unclassified	3	5	1	7
REGION				
Northeast	768	780	597	951
Southeast	874	974	581	1267
Central	1017	1080	640	1457
West	1212	1264	816	1660
PARENTS' EDUCATION				
Less than high school	165	175	93	247
High school	594	638	392	840
Greater than high school	186	219	129	276
Graduated college	1570	1727	1153	2144
Unknown	1342	1324	858	1808
TYPE OF LOCATION				
Large city	585	663	453	795
Midsize city	772	765	483	1054
Urban fringe of large city	990	993	705	1278
Urban fringe of midsize city	718	797	496	1019
Large town	0	0	0	0
Small town	435	482	269	648
Rural/MSA	23	29	12	40
Rural/NonMSA	348	369	216	501

****Note:*** *NAEP reporting subgroups are described in Appendix B.*

Table 19-19

Number of Students in the Reading and Writing Long-Term Trend Sample by Type of Eligibility and Subgroup Classification, Age 13/Grade 8*

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
TOTAL	3989	4058	2500	5547
SEX				
Male	2006	2077	1192	2891
Female	1983	1981	1308	2656
RACE/ETHNICITY				
White	2709	2821	1745	3785
Black	529	502	281	750
Hispanic	511	504	324	691
Asian American	141	146	98	189
American Indian	91	79	46	124
Unclassified	8	6	6	8
REGION				
Northeast	843	859	601	1101
Southeast	940	956	545	1351
Central	934	927	550	1311
West	1272	1316	804	1784
PARENTS' EDUCATION				
Less than high school	285	290	148	427
High school	1054	1062	604	1512
Greater than high school	417	475	296	596
Graduated college	1860	1881	1253	2488
Unknown	348	331	189	490
TYPE OF LOCATION				
Large city	474	490	303	661
Midsize city	814	773	449	1138
Urban fringe of large city	879	905	606	1178
Urban fringe of midsize city	881	918	571	1228
Large town	0	0	0	0
Small town	468	496	273	691
Rural/MSA	96	94	62	128
Rural/NonMSA	377	382	236	523

****Note:*** *NAEP reporting subgroups are described in Appendix B.*

Table 19-20

Number of Students in the Reading and Writing Long-Term Trend Sample by Type of Eligibility and Subgroup Classification, Age 17/Grade 11*

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
TOTAL	3811	3948	2919	4840
SEX				
Male	1871	1899	1370	2400
Female	1940	2049	1549	2440
RACE/ETHNICITY				
White	2622	2665	2094	3193
Black	563	642	396	809
Hispanic	418	427	293	552
Asian American	139	150	91	198
American Indian	64	57	40	81
Unclassified	5	7	5	7
REGION				
Northeast	840	878	622	1096
Southeast	928	966	698	1196
Central	932	942	726	1148
West	1111	1162	873	1400
PARENTS' EDUCATION				
Less than high school	275	284	171	388
High school	1010	1027	734	1303
Greater than high school	727	768	572	923
Graduated college	1675	1750	1368	2057
Unknown	113	108	65	156
TYPE OF LOCATION				
Large city	510	574	366	718
Midsize city	631	634	488	777
Urban fringe of large city	838	850	629	1059
Urban fringe of midsize city	889	910	704	1095
Large town	0	0	0	0
Small town	510	539	390	659
Rural/MSA	59	66	49	76
Rural/NonMSA	374	375	293	456

****Note****: NAEP reporting subgroups are described in Appendix B.*

Table 19-21

Number of Students in the Mathematics and Science Long-Term Trend Sample by Type of Eligibility and Subgroup Classification, Age 9/Grade 4*

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
TOTAL	5663	3751	3751	5663
SEX				
Male	2755	1746	1746	2755
Female	2908	2005	2005	2908
RACE/ETHNICITY				
White	3591	2375	2375	3591
Black	724	522	522	724
Hispanic	1106	689	689	1106
Asian American	103	84	84	103
American Indian	128	75	75	128
Unclassified	11	6	6	11
REGION				
Northeast	1107	829	829	1107
Southeast	1287	853	853	1287
Central	1539	931	931	1539
West	1730	1138	1138	1730
PARENTS' EDUCATION				
Less than high school	235	136	136	235
High school	806	519	519	806
Greater than high school	381	274	274	381
Graduated college	2447	1715	1715	2447
Unknown	1782	1103	1103	1782
TYPE OF LOCATION				
Large city	779	599	599	779
Midsize city	1109	699	699	1109
Urban fringe of large city	1413	957	957	1413
Urban fringe of midsize city	1071	742	742	1071
Large town	0	0	0	0
Small town	720	417	417	720
Rural/MSA	37	23	23	37
Rural/NonMSA	534	314	314	534

**Note: NAEP reporting subgroups are described in Appendix B.*

Table 19-22

Number of Students in the Mathematics and Science Long-Term Trend Sample by Type of Eligibility and Subgroup Classification, Age 13/Grade 8*

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
TOTAL	6052	3824	3824	6052
SEX				
Male	2994	1743	1743	2994
Female	3058	2081	2081	3058
RACE/ETHNICITY				
White	4165	2713	2713	4165
Black	765	446	446	765
Hispanic	799	470	470	799
Asian American	214	148	148	214
American Indian	86	32	32	86
Unclassified	23	15	15	23
REGION				
Northeast	1193	851	851	1193
Southeast	1440	886	886	1440
Central	1394	832	832	1394
West	2025	1255	1255	2025
PARENTS' EDUCATION				
Less than high school	393	218	218	393
High school	1367	823	823	1367
Greater than high school	1013	672	672	1013
Graduated college	2722	1834	1834	2722
Unknown	521	262	262	521
TYPE OF LOCATION				
Large city	774	521	521	774
Midsize city	1113	659	659	1113
Urban fringe of large city	1361	943	943	1361
Urban fringe of midsize city	1260	808	808	1260
Large town	0	0	0	0
Small town	795	482	482	795
Rural/MSA	174	107	107	174
Rural/NonMSA	575	304	304	575

****Note:*** *NAEP reporting subgroups are described in Appendix B.*

Table 19-23

Number of Students in the Mathematics and Science Long-Term Trend Sample by Type of Eligibility and Subgroup Classification, Age 17/Grade 11*

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
TOTAL	3813	2832	2832	3813
SEX				
Male	1862	1328	1328	1862
Female	1951	1504	1504	1951
RACE/ETHNICITY				
White	2651	2056	2056	2651
Black	575	375	375	575
Hispanic	394	267	267	394
Asian American	134	99	99	134
American Indian	41	24	24	41
Unclassified	18	11	11	18
REGION				
Northeast	911	669	669	911
Southeast	961	700	700	961
Central	920	692	692	920
West	1021	771	771	1021
PARENTS' EDUCATION				
Less than high school	274	171	171	274
High school	854	582	582	854
Greater than high school	924	727	727	924
Graduated college	1635	1280	1280	1635
Unknown	115	66	66	115
TYPE OF LOCATION				
Large city	641	447	447	641
Midsize city	521	396	396	521
Urban fringe of large city	765	555	555	765
Urban fringe of midsize city	818	614	614	818
Large town	0	0	0	0
Small town	590	461	461	590
Rural/MSA	75	56	56	75
Rural/NonMSA	403	303	303	403

**Note: NAEP reporting subgroups are described in Appendix B.*

Table 19-24

Number of Excluded Students in Main Samples
by Type of Eligibility and Subgroup Classification
Age 9/Grade 4

	Eligible by			
	Age	Grade	Age & Grade	Total
TOTAL	955	1025	471	1512
SEX				
Male	589	660	286	965
Female	366	365	185	547
RACE/ETHNICITY				
White	384	381	161	606
Black	214	266	101	379
Hispanic	302	329	179	452
Other	53	48	29	73
REGION				
Northeast	189	204	100	294
Southeast	192	233	80	345
Central	172	183	74	281
West	402	405	217	592
REASON FOR EXCLUSION				
Disability	547	637	250	936
Limited English Proficiency	340	330	204	467
Both Disability and LEP	30	30	9	51
Nonreader	9	5	1	13
Other	29	22	7	44

Table 19-25

Number of Excluded Students in Main Samples
by Type of Eligibility and Subgroup Classification
Age 13/Grade 9

	Eligible by			
	Age	Grade	Age & Grade	Total
TOTAL	1119	1278	358	2044
SEX				
Male	760	849	224	1387
Female	359	429	134	657
RACE/ETHNICITY				
White	524	580	160	946
Black	274	289	65	500
Hispanic	275	344	116	503
Other	45	61	16	91
REGION				
Northeast	223	223	76	372
Southeast	300	353	73	580
Central	226	283	56	455
West	370	419	153	637
REASON FOR EXCLUSION				
Disability	832	915	241	1508
Limited English Proficiency	226	301	106	424
Both Disability and LEP	18	21	3	36
Nonreader	3	4	1	6
Other	35	31	6	60

Table 19-26

Number of Excluded Students in Main Samples
by Type of Eligibility and Subgroup Classification
Age 17/Grade 12

	Eligible by			
	Age	Grade	Age & Grade	Total
TOTAL	1190	948	267	1874
SEX				
Male	775	580	166	1191
Female	415	368	101	683
RACE/ETHNICITY				
White	562	530	155	937
Black	235	194	30	401
Hispanic	330	170	67	434
Other	60	51	14	97
REGION				
Northeast	380	249	102	528
Southeast	272	259	46	487
Central	197	182	44	335
West	341	258	75	524
REASON FOR EXCLUSION				
Disability	840	732	189	1384
Limited English Proficiency	306	169	69	407
Both Disability and LEP	22	22	4	41
Nonreader	3	3	1	5
Other	16	16	2	30

Table 19-27

Number of Excluded Students in Long-Term Trend Samples by Type of Eligibility and Subgroup Classification

Age 9/Grade 4

	Eligible by			
	Age	Grade	Age & Grade	Total
TOTAL	796	535	347	984
SEX				
Male	488	330	204	614
Female	308	205	143	370
RACE/ETHNICITY				
White	337	220	114	443
Black	124	82	47	159
Hispanic	272	196	155	313
Other	63	37	31	69
REGION				
Northeast	130	93	68	155
Southeast	164	97	47	214
Central	129	84	49	164
West	373	261	183	451
REASON FOR EXCLUSION				
Disability	429	278	154	553
Limited English Proficiency	257	188	155	290
Both Disability and LEP	15	15	7	23
Nonreader	48	27	17	58
Other	47	27	14	60

Table 19-28

Number of Excluded Students in Long-Term Trend Samples by Type of Eligibility and Subgroup Classification

Age 13/Grade 8

	Eligible by			
	Age	Grade	Age & Grade	Total
TOTAL	639	365	198	806
SEX				
Male	403	224	116	511
Female	236	141	82	295
RACE/ETHNICITY				
White	302	167	77	392
Black	121	54	26	149
Hispanic	169	112	70	211
Other	47	32	25	54
REGION				
Northeast	101	47	30	118
Southeast	150	73	30	193
Central	139	75	31	183
West	249	170	107	312
REASON FOR EXCLUSION				
Disability	420	226	101	545
Limited English Proficiency	171	114	82	203
Both Disability and LEP	21	14	10	25
Nonreader	15	6	4	17
Other	10	2	1	11

Table 19-29

Number of Excluded Students in Long-Term Trend Samples by Type of Eligibility and Subgroup Classification

Age 17/Grade 12

	Eligible by			
	Age	Grade	Age & Grade	Total
TOTAL	388	274	158	504
SEX				
Male	248	171	95	324
Female	140	103	63	180
RACE/ETHNICITY				
White	218	149	91	276
Black	68	58	26	100
Hispanic	77	52	33	96
Other	25	15	8	32
REGION				
Northeast	109	61	36	134
Southeast	85	72	34	123
Central	77	42	27	92
West	117	99	61	155
REASON FOR EXCLUSION				
Disability	273	192	109	356
Limited English Proficiency	81	55	34	102
Both Disability and LEP	7	1	0	8
Nonreader	8	7	5	10
Other	19	19	10	28

Table 19-30

Number of Students by Sample and Age Class

	—— Age Class 9 ——		—— Age Class 13 ——		—— Age Class 17 ——	
	Total	Sum of Weights	Total	Sum of Weights	Total	Sum of Weights
MAIN ASSESSMENTS						
Reading	9680	4990698	20873	5002736	19520	3790666
U.S. History	7205	4991663	11669	5001435	10837	3748557
Geography	7205	4994407	9186	4997641	8589	3742046
LONG-TERM TREND ASSESSMENTS						
Reading and Writing	5335	4885228	5547	4905112	4840	4251875
Mathematics and Science	5663	3308719	6052	3353585	3813	3078721
EXCLUDED STUDENTS						
Main	1512	339858	2044	289332	1874	176182
Long-Term Trend	984	446460	806	382660	504	278464

Table 19-31

Number of Students Assessed in the Reading Main Sample
Age 9/Grade 4

		Grade		
	< 4	= 4	> 4	Total
AGE < 9				
Unweighted N	0	37	0	37
Estimated population size	0	18877	0	18877
Standard error	0	3863	0	3863
Coefficient of variation	0.00	20.46	0.00	20.46
AGE = 9				
Unweighted N	2288	4744	10	7042
Estimated population size	1457321	2114256	5967	3577543
Standard error	10490	37327	2113	39157
Coefficient of variation	0.72	1.77	35.41	1.09
AGE > 9				
Unweighted N	0	2601	0	2601
Estimated population size	0	1394278	0	1394278
Standard error	0	37280	0	37280
Coefficient of variation	0.00	2.67	0.00	2.67
AGE TOTAL				
Unweighted N	2288	7382	10	9680
Estimated population size	1457321	3527411	5967	4990699
Standard error	10490	14975	2113	20910
Coefficient of variation	0.72	0.42	35.41	0.42

***Note**: *The coefficient of variation of the estimated population is defined as 100 times its standard error divided by the estimated population size.*

Table 19-32

Number of Students Assessed in the U.S. History Main Sample
Age 9/Grade 4

	Grade			
	< 4	= 4	> 4	Total
AGE < 9				
Unweighted N	0	25	0	25
Estimated population size	0	17100	0	17100
Standard error	0	3916	0	3916
Coefficient of variation	0.00	22.90	0.00	22.90
AGE = 9				
Unweighted N	1691	3533	15	5239
Estimated population size	1454975	2088451	8895	3552320
Standard error	11799	34739	2583	36283
Coefficient of variation	0.81	1.66	29.04	1.02
AGE > 9				
Unweighted N	0	1941	0	1941
Estimated population size	0	1422243	0	1422243
Standard error	0	37044	0	37044
Coefficient of variation	0.00	2.60	0.00	2.60
AGE TOTAL				
Unweighted N	1691	5499	15	7205
Estimated population size	1454975	3527794	8895	4991663
Standard error	11799	14377	2583	20678
Coefficient of variation	0.81	0.41	29.04	0.41

*Note: The coefficient of variation of the estimated population is defined as 100 times its standard error divided by the estimated population size.

Table 19-33

Number of Students Assessed in the Geography Main Sample
Age 9/Grade 4

	Grade			
	< 4	= 4	> 4	Total
AGE < 9				
Unweighted N	0	16	0	16
Estimated population size	0	9929	0	9929
Standard error	0	2923	0	2923
Coefficient of variation	0.00	29.44	0.00	29.44
AGE = 9				
Unweighted N	1686	3469	12	5167
Estimated population size	1455387	2073294	8205	3536886
Standard error	10522	38703	2247	40820
Coefficient of variation	0.72	1.87	27.39	1.15
AGE > 9				
Unweighted N	0	2022	0	2022
Estimated population size	0	1447592	0	1447592
Standard error	0	38736	0	38736
Coefficient of variation	0.00	2.68	0.00	2.68
AGE TOTAL				
Unweighted N	1686	5507	12	7205
Estimated population size	1455387	3530816	8205	4994408
Standard error	10522	13584	2247	19390
Coefficient of variation	0.72	0.38	27.39	0.39

***Note**: The coefficient of variation of the estimated population is defined as 100 times its standard error divided by the estimated population size.*

BEST COPY AVAILABLE

Table 19-34

Number of Students Assessed in the Reading Main Sample
Age 13/Grade 8

	Grade			
	< 8	= 8	> 8	Total
AGE < 13				
Unweighted N	0	100	0	100
Estimated population size	0	23677	0	23677
Standard error	0	2660	0	2660
Coefficient of variation	0.00	11.23	0.00	11.23
AGE = 13				
Unweighted N	5232	9564	35	14831
Estimated population size	1538330	1963651	15783	3517764
Standard error	11596	26930	7205	29050
Coefficient of variation	0.75	1.37	45.65	0.83
AGE > 13				
Unweighted N	0	5942	0	5942
Estimated population size	0	1461296	0	1461296
Standard error	0	26768	0	26768
Coefficient of variation	0.00	1.83	0.00	1.83
AGE TOTAL				
Unweighted N	5232	15606	35	20873
Estimated population size	1538330	3448624	15783	5002737
Standard error	11596	11901	7205	19703
Coefficient of variation	0.75	0.35	45.65	0.39

*Note: The coefficient of variation of the estimated population is defined as 100 times its standard error divided by the estimated population size.

Table 19-35

Number of Students Assessed in the U.S. History Main Sample
Age 13/Grade 8

	Grade			
	< 8	= 8	> 8	Total
AGE < 13				
Unweighted N	0	50	0	50
Estimated population size	0	20029	0	20029
Standard error	0	3633	0	3633
Coefficient of variation	0.00	18.14	0.00	18.14
AGE = 13				
Unweighted N	2889	5400	13	8302
Estimated population size	1533200	1973461	19043	3525703
Standard error	15124	27355	8974	30010
Coefficient of variation	0.99	1.39	47.12	0.85
AGE > 13				
Unweighted N	0	3317	0	3317
Estimated population size	0	1455703	0	1455703
Standard error	0	28435	0	28435
Coefficient of variation	0.00	1.95	0.00	1.95
AGE TOTAL				
Unweighted N	2889	8767	13	11669
Estimated population size	1533200	3449193	19043	5001435
Standard error	15124	10765	8974	20105
Coefficient of variation	0.99	0.31	47.12	0.40

***Note**: *The coefficient of variation of the estimated population is defined as 100 times its standard error divided by the estimated population size.*

Table 19-36

Number of Students Assessed in the Geography Main Sample
Age 13/Grade 8

	Grade			
	< 8	= 8	> 8	Total
AGE < 13				
Unweighted N	0	33	0	33
Estimated population size	0	16804	0	16804
Standard error	0	4066	0	4066
Coefficient of variation	0.00	24.20	0.00	24.20
AGE = 13				
Unweighted N	2299	4250	9	6558
Estimated population size	1535205	1986049	15292	3536546
Standard error	14287	32277	7477	35724
Coefficient of variation	0.93	1.63	48.89	1.01
AGE > 13				
Unweighted N	0	2595	0	2595
Estimated population size	0	1444292	0	1444292
Standard error	0	33525	0	33525
Coefficient of variation	0.00	2.32	0.00	2.32
AGE TOTAL				
Unweighted N	2299	6878	9	9186
Estimated population size	1535205	3447145	15292	4997642
Standard error	14287	11457	7477	20771
Coefficient of variation	0.93	0.33	48.89	0.42

Note: *The coefficient of variation of the estimated population is defined as 100 times its standard error divided by the estimated population size.*

Table 19-37

Number of Students Assessed in the Reading Main Sample
Age 17/Grade 12

	Grade			
	< 12	= 12	> 12	Total
AGE < 17				
Unweighted N	0	151	0	151
Estimated population size	0	21821	0	21821
Standard error	0	2309	0	2309
Coefficient of variation	0.00	10.58	0.00	10.58
AGE = 17				
Unweighted N	5339	9329	0	14668
Estimated population size	1200983	1654993	0	2855976
Standard error	8535	10080	0	15091
Coefficient of variation	0.71	0.61	0.00	0.53
AGE > 17				
Unweighted N	0	4701	0	4701
Estimated population size	0	912871	0	912871
Standard error	0	23965	0	23965
Coefficient of variation	0.00	2.63	0.00	2.63
AGE TOTAL				
Unweighted N	5339	14181	0	19520
Estimated population size	1200983	2589684	0	3790667
Standard error	8535	20129	0	24249
Coefficient of variation	0.71	0.78	0.00	0.64

***Note:** *The coefficient of variation of the estimated population is defined as 100 times its standard error divided by the estimated population size.*

Table 19-38

Number of Students Assessed in the U.S. History Main Sample
Age 17/Grade 12

	Grade			
	< 12	= 12	> 12	Total
AGE < 17				
Unweighted N	0	101	0	101
Estimated population size	0	25057	0	25057
Standard error	0	3420	0	3420
Coefficient of variation	0.00	13.65	0.00	13.65
AGE = 17				
Unweighted N	3019	5113	0	8132
Estimated population size	1202660	1638951	0	2841610
Standard error	8341	15272	0	16820
Coefficient of variation	0.69	0.93	0.00	0.59
AGE > 17				
Unweighted N	0	2604	0	2604
Estimated population size	0	881891	0	881891
Standard error	0	34107	0	34107
Coefficient of variation	0.00	3.87	0.00	3.87
AGE TOTAL				
Unweighted N	3019	7818	0	10837
Estimated population size	1202660	2545898	0	3748558
Standard error	8341	23494	0	27139
Coefficient of variation	0.69	0.92	0.00	0.72

***Note**: The coefficient of variation of the estimated population is defined as 100 times its standard error divided by the estimated population size.*

Table 19-39

Number of Students Assessed in the Geography Main Sample
Age 17/Grade 12

	Grade			
	< 12	= 12	> 12	Total
AGE < 17				
Unweighted N	0	89	0	89
Estimated population size	0	28910	0	28910
Standard error	0	4155	0	4155
Coefficient of variation	0.00	14.37	0.00	14.37
AGE = 17				
Unweighted N	2355	4118	0	6473
Estimated population size	1199733	1641129	0	2840862
Standard error	8660	14771	0	18038
Coefficient of variation	0.72	0.90	0.00	0.63
AGE > 17				
Unweighted N	0	2027	0	2027
Estimated population size	0	872275	0	872275
Standard error	0	28855	0	28855
Coefficient of variation	0.00	3.31	0.00	3.31
AGE TOTAL				
Unweighted N	2355	6234	0	8589
Estimated population size	1199733	2542314	0	3742047
Standard error	8660	23213	0	27807
Coefficient of variation	0.72	0.91	0.00	0.74

***Note:** The coefficient of variation of the estimated population is defined as 100 times its standard error divided by the estimated population size.*

Table 19-40

Number of Students Assessed in the Reading and Writing Long-Term Trend Sample
Age 9/Grade 4

	Grade			
	< 4	= 4	> 4	Total
AGE < 9				
Unweighted N	0	21	0	21
Estimated population size	0	22464	0	22464
Standard error	0	5200	0	5200
Coefficient of variation	0.00	23.15	0.00	23.15
AGE = 9				
Unweighted N	1226	2634	11	3871
Estimated population size	1391849	2088006	14346	3494202
Standard error	15580	42480	4690	49419
Coefficient of variation	1.12	2.03	32.69	1.41
AGE > 9				
Unweighted N	0	1443	0	1443
Estimated population size	0	1368563	0	1368563
Standard error	0	35461	0	35461
Coefficient of variation	0.00	2.59	0.00	2.59
AGE TOTAL				
Unweighted N	1226	4098	11	5335
Estimated population size	1391849	3479033	14346	4885228
Standard error	15580	21423	4690	31200
Coefficient of variation	1.12	0.62	32.69	0.64

***Note**: *The coefficient of variation of the estimated population is defined as 100 times its standard error divided by the estimated population size.*

Table 19-41

Number of Students Assessed in the Reading and Writing Long-Term Trend Sample
Age 13/Grade 8

	Grade			
	< 8	= 8	> 8	Total
AGE < 13				
Unweighted N	0	21	0	21
Estimated population size	0	20616	0	20616
Standard error	0	5104	0	5104
Coefficient of variation	0.00	24.76	0.00	24.76
AGE = 13				
Unweighted N	1489	2500	0	3989
Estimated population size	1503678	1933571	0	3437249
Standard error	17852	34166	0	37986
Coefficient of variation	1.19	1.77	0.00	1.11
AGE > 13				
Unweighted N	0	1537	0	1537
Estimated population size	0	1447247	0	1447247
Standard error	0	37725	0	37725
Coefficient of variation	0.00	2.61	0.00	2.61
AGE TOTAL				
Unweighted N	1489	4058	0	5547
Estimated population size	1503678	3401434	0	4905112
Standard error	17852	20235	0	30473
Coefficient of variation	1.19	0.59	0.00	0.62

***Note:** The coefficient of variation of the estimated population is defined as 100 times its standard error divided by the estimated population size.*

Table 19-42

Number of Students Assessed in the Reading and Writing Long-Term Trend Sample
Age 17/Grade 11

	Grade			
	< 11	= 11	> 11	Total
AGE < 17				
Unweighted N	0	329	0	329
Estimated population size	0	335803	0	335803
Standard error	0	28190	0	28190
Coefficient of variation	0.00	8.39	0.00	8.39
AGE = 17				
Unweighted N	706	2919	186	3811
Estimated population size	915574	1977526	231612	3124712
Standard error	24792	4876	25190	15245
Coefficient of variation	2.71	0.25	10.88	0.49
AGE > 17				
Unweighted N	0	700	0	700
Estimated population size	0	791360	0	791360
Standard error	0	27251	0	27251
Coefficient of variation	0.00	3.44	0.00	3.44
AGE TOTAL				
Unweighted N	706	3948	186	4840
Estimated population size	915574	3104690	231612	4251875
Standard error	24792	14958	25190	24031
Coefficient of variation	2.71	0.48	10.88	0.57

***Note:** *The coefficient of variation of the estimated population is defined as 100 times its standard error divided by the estimated population size.*

Table 19-43

Number of Students Assessed in the Mathematics and Science Long-Term Trend Sample

	Grade			
	< 4	= 4	> 4	Total
AGE 9				
Unweighted N	1901	3751	11	5663
Estimated population size	1104497	2196785	7438	3308720
Standard error	42566	48691	3113	27093
Coefficient of variation	3.85	2.22	41.85	0.82

	Grade			
	< 8	= 8	> 8	Total
AGE 13				
Unweighted N	2214	3824	14	6052
Estimated population size	1262502	2063833	27250	3353585
Standard error	43352	48003	23676	27789
Coefficient of variation	3.43	2.33	86.88	0.83

	Grade			
	< 11	= 11	> 11	Total
AGE 17				
Unweighted N	728	2832	253	3813
Estimated population size	634977	2249987	193757	3078722
Standard error	48637	52630	19360	16549
Coefficient of variation	7.66	2.34	9.99	0.54

***Note**: The coefficient of variation of the estimated population is defined as 100 times its standard error divided by the estimated population size.*

Table 19-44

Weighted Percentage of Students in the Reading Main Sample by Type of Eligibility and Subgroup Classification, Age 9/Grade 4*

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
SEX				
Male	51.64	51.24	47.60	53.07
Female	48.36	48.76	52.40	46.93
RACE/ETHNICITY				
White	69.54	68.73	69.33	69.05
Black	14.69	15.00	14.83	14.85
Hispanic	10.71	11.58	10.32	11.49
Asian American	1.94	1.89	2.50	1.67
Pacific Islander	0.74	0.70	0.78	0.69
American Indian	1.78	1.55	1.53	1.72
Unclassified	0.60	0.56	0.71	0.53
REGION				
Northeast	23.22	22.71	27.15	21.19
Southeast	23.48	23.42	21.49	24.28
Central	25.15	24.67	21.89	26.19
West	28.15	29.20	29.46	28.34
PARENTS' EDUCATION				
Less than high school	4.14	4.14	3.55	4.39
High school	11.90	12.54	11.60	12.48
Greater than high school	6.73	7.57	7.72	6.91
Graduated college	41.23	41.91	44.05	40.52
Unknown	35.73	33.61	32.94	35.41
TYPE OF LOCATION				
Large city	9.66	11.23	12.04	9.76
Midsize city	23.72	23.95	23.67	23.91
Urban fringe of large city	27.06	24.51	26.77	25.38
Urban fringe of midsize city	16.62	15.71	16.43	16.06
Large town	2.94	3.19	3.28	2.97
Small town	10.39	11.20	9.49	11.35
Rural/MSA	1.68	2.08	1.69	1.96
Rural/NonMSA	7.93	8.13	6.63	8.62
ESTIMATED TOTAL POPULATION	3577543	3527410	2114255	4990698

****Note***: *NAEP reporting subgroups are described in Appendix B.*

Table 19-45

Weighted Percentage of Students in the U.S. History Main Sample by Type of Eligibility and Subgroup Classification, Age 9/Grade 4*

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
SEX				
Male	49.63	49.52	45.67	51.21
Female	50.37	50.48	54.33	48.79
RACE/ETHNICITY				
White	69.40	68.75	69.11	69.06
Black	14.67	15.08	14.72	14.93
Hispanic	10.90	11.50	10.69	11.41
Asian American	1.79	1.66	2.14	1.55
Pacific Islander	0.96	0.93	1.08	0.89
American Indian	1.69	1.60	1.50	1.70
Unclassified	0.58	0.50	0.75	0.45
REGION				
Northeast	22.60	21.76	26.42	20.41
Southeast	23.13	23.21	21.21	23.99
Central	25.99	25.16	22.97	26.67
West	28.29	29.86	29.41	28.93
PARENTS' EDUCATION				
Less than high school	3.59	3.52	2.69	3.92
High school	11.00	12.14	10.92	11.84
Greater than high school	6.99	7.77	8.01	7.12
Graduated college	40.80	42.50	44.09	40.62
Unknown	37.22	33.83	34.01	36.17
TYPE OF LOCATION				
Large city	10.88	11.29	12.87	10.33
Midsize city	23.86	23.71	23.29	23.99
Urban fringe of large city	25.82	24.64	27.49	24.29
Urban fringe of midsize city	15.65	15.41	14.71	15.87
Large town	2.79	3.02	3.01	2.86
Small town	10.81	10.89	9.52	11.41
Rural/MSA	2.11	2.80	2.09	2.61
Rural/NonMSA	8.09	8.24	7.01	8.64
ESTIMATED TOTAL POPULATION	3552320	3527793	2088450	4991663

**Note: NAEP reporting subgroups are described in Appendix B.*

Table 19-46

Weighted Percentage of Students in the Geography Main Sample
by Type of Eligibility and Subgroup Classification, Age 9/Grade 4*

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
SEX				
Male	49.39	50.60	46.45	51.47
Female	50.61	49.40	53.55	48.53
RACE/ETHNICITY				
White	69.27	68.69	68.91	69.01
Black	14.77	15.08	14.85	14.96
Hispanic	10.79	11.55	10.52	11.44
Asian American	1.70	1.73	2.18	1.52
Pacific Islander	1.05	0.94	1.24	0.89
American Indian	1.61	1.49	1.62	1.53
Unclassified	0.81	0.52	0.68	0.66
REGION				
Northeast	22.61	21.74	26.55	20.37
Southeast	23.45	23.04	21.45	23.99
Central	25.80	25.48	22.82	26.81
West	28.14	29.74	29.19	28.83
PARENTS' EDUCATION				
Less than high school	3.78	3.96	3.34	4.09
High school	11.15	11.98	10.36	12.06
Greater than high school	6.15	7.33	7.20	6.55
Graduated college	40.87	42.08	44.19	40.35
Unknown	37.47	34.32	34.70	36.39
TYPE OF LOCATION				
Large city	10.38	11.23	12.79	9.98
Midsize city	23.90	24.04	23.38	24.22
Urban fringe of large city	26.01	24.31	26.28	24.70
Urban fringe of midsize city	15.80	15.51	15.35	15.79
Large town	2.92	3.14	3.12	2.99
Small town	10.52	10.85	9.49	11.18
Rural/MSA	2.51	2.79	2.54	2.69
Rural/NonMSA	7.96	8.11	7.06	8.44
ESTIMATED TOTAL POPULATION	3536886	3530815	2073293	4994407

**Note*: *NAEP reporting subgroups are described in Appendix B.*

Table 19-47

Weighted Percentage of Students in the Reading Main Sample by Type of Eligibility and Subgroup Classification, Age 13/Grade 8*

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
SEX				
Male	51.09	50.41	45.63	52.76
Female	48.91	49.59	54.37	47.24
RACE/ETHNICITY				
White	68.70	69.14	71.46	67.93
Black	14.93	15.20	12.85	15.94
Hispanic	11.17	10.80	9.97	11.39
Asian American	1.88	1.84	2.40	1.65
Pacific Islander	1.27	1.30	1.67	1.14
American Indian	1.69	1.35	1.22	1.64
Unclassified	0.35	0.37	0.44	0.33
REGION				
Northeast	20.66	20.46	23.97	19.22
Southeast	23.98	25.65	21.99	25.91
Central	24.74	24.09	22.71	25.09
West	30.62	29.80	31.34	29.77
PARENTS' EDUCATION				
Less than high school	6.74	7.02	5.33	7.48
High school	21.85	21.57	20.37	22.24
Greater than high school	17.87	19.21	19.93	17.99
Graduated college	42.21	42.65	46.21	40.94
Unknown	10.87	9.15	7.79	10.89
TYPE OF LOCATION				
Large city	12.74	12.62	13.65	12.30
Midsize city	22.09	21.68	21.83	21.92
Urban fringe of large city	19.94	19.93	21.84	19.19
Urban fringe of midsize city	16.51	17.63	18.05	16.68
Large town	1.96	2.01	2.02	1.98
Small town	11.23	11.71	10.02	12.04
Rural/MSA	5.08	4.84	4.16	5.28
Rural/NonMSA	10.43	9.56	8.44	10.62
ESTIMATED TOTAL POPULATION	3517763	3448623	1963650	5002736

**Note: NAEP reporting subgroups are described in Appendix B.*

Table 19-48

Weighted Percentage of Students in the U.S. History Main Sample by Type of Eligibility and Subgroup Classification, Age 13/Grade 8*

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
SEX				
Male	49.55	49.99	44.59	51.81
Female	50.45	50.01	55.41	48.19
RACE/ETHNICITY				
White	68.62	69.03	71.40	67.81
Black	15.01	15.27	12.84	16.05
Hispanic	11.40	10.82	10.46	11.37
Asian American	1.97	1.95	2.44	1.77
Pacific Islander	1.13	1.17	1.43	1.04
American Indian	1.61	1.41	1.08	1.69
Unclassified	0.24	0.35	0.36	0.27
REGION				
Northeast	21.30	20.07	23.55	19.57
Southeast	23.66	25.47	22.51	25.36
Central	24.32	23.79	22.26	24.77
West	30.72	30.66	31.67	30.30
PARENTS' EDUCATION				
Less than high school	7.05	7.11	5.49	7.71
High school	22.13	22.62	20.42	23.14
Greater than high school	17.47	18.77	19.19	17.68
Graduated college	41.91	41.64	46.16	40.05
Unknown	11.08	9.44	8.38	11.02
TYPE OF LOCATION				
Large city	12.38	12.65	13.09	12.28
Midsize city	22.81	23.69	23.47	23.16
Urban fringe of large city	20.96	18.95	21.53	19.35
Urban fringe of midsize city	16.05	16.90	17.47	16.07
Large town	1.82	2.16	1.91	2.02
Small town	11.69	12.19	11.02	12.30
Rural/MSA	4.99	4.69	4.03	5.16
Rural/NonMSA	9.29	8.78	7.48	9.66
ESTIMATED TOTAL POPULATION	3525703	3449192	1973460	5001435

**__Note__: NAEP reporting subgroups are described in Appendix B.*

Table 13.49

Weighted Percentage of Students in the Geography Main Sample by Type of Eligibility and Subgroup Classification, Age 13/Grade 8*

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
SEX				
Male	49.50	50.74	45.13	52.09
Female	50.50	49.26	54.87	47.91
RACE/ETHNICITY				
White	68.64	69.11	71.42	67.86
Black	15.02	15.25	12.78	16.07
Hispanic	11.27	10.78	10.31	11.31
Asian American	1.88	1.78	2.35	1.63
Pacific Islander	1.33	1.11	1.48	1.12
American Indian	1.60	1.53	1.31	1.67
Unclassified	0.26	0.45	0.36	0.35
REGION				
Northeast	21.71	20.14	24.29	19.60
Southeast	23.35	25.11	21.67	25.23
Central	23.75	24.09	21.68	24.81
West	31.18	30.66	32.36	30.36
PARENTS' EDUCATION				
Less than high school	6.70	6.99	5.61	7.33
High school	21.61	22.14	19.48	22.82
Greater than high school	17.55	19.24	20.33	17.61
Graduated college	42.21	41.75	45.61	40.54
Unknown	11.57	9.50	8.61	11.32
TYPE OF LOCATION				
Large city	12.94	12.96	13.89	12.58
Midsize city	23.68	23.01	23.71	23.21
Urban fringe of large city	20.76	19.27	21.63	19.39
Urban fringe of midsize city	15.37	17.16	16.56	16.13
Large town	1.98	1.95	2.18	1.88
Small town	11.39	11.83	10.76	11.94
Rural/MSA	4.81	5.08	4.11	5.28
Rural/NonMSA	9.07	8.74	7.16	9.60
ESTIMATED TOTAL POPULATION	3536545	3447144	1986048	4997641

**Note: NAEP reporting subgroups are described in Appendix B.*

Table 19-50

Weighted Percentage of Students in the Reading Main Sample by Type of Eligibility and Subgroup Classification, Age 17/Grade 12*

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
SEX				
Male	50.11	50.21	45.77	52.08
Female	49.89	49.79	54.23	47.92
RACE/ETHNICITY				
White	70.13	73.70	75.97	70.01
Black	14.85	12.63	11.17	14.94
Hispanic	10.11	8.41	7.69	10.01
Asian American	2.47	2.82	2.75	2.59
Pacific Islander	1.20	1.10	1.36	1.06
American Indian	0.94	1.00	0.79	1.04
Unclassified	0.30	0.34	0.26	0.35
REGION				
Northeast	19.51	20.46	22.78	18.73
Southeast	24.41	23.26	22.45	24.48
Central	26.84	26.99	26.21	27.22
West	29.25	29.29	28.56	29.58
PARENTS' EDUCATION				
Less than high school	7.80	7.30	5.56	8.44
High school	21.28	21.11	19.17	22.09
Greater than high school	24.92	25.10	26.15	24.50
Graduated college	42.53	43.55	46.80	41.36
Unknown	2.94	2.48	1.87	3.09
TYPE OF LOCATION				
Large city	10.44	9.59	9.63	10.21
Midsize city	22.36	21.87	22.74	21.86
Urban fringe of large city	24.88	25.99	27.45	24.52
Urban fringe of midsize city	14.18	14.29	14.38	14.17
Large town	1.91	2.01	2.04	1.92
Small town	12.82	12.35	11.38	13.12
Rural/MSA	4.64	4.74	4.22	4.89
Rural/NonMSA	8.78	9.17	8.18	9.31
ESTIMATED TOTAL POPULATION	2855975	2589683	1654992	3790666

**Note*: *NAEP reporting subgroups are described in Appendix B.*

Table 19-51

Weighted Percentage of Students in the U.S. History Main Sample by Type of Eligibility and Subgroup Classification, Age 17/Grade 12*

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
SEX				
Male	51.39	50.21	46.21	52.85
Female	48.61	49.79	53.79	47.15
RACE/ETHNICITY				
White	70.34	73.68	76.39	69.96
Black	14.88	12.41	11.20	14.81
Hispanic	9.88	8.79	7.28	10.28
Asian American	2.59	2.86	2.77	2.69
Pacific Islander	0.98	1.05	1.22	0.92
American Indian	0.94	0.83	0.79	0.94
Unclassified	0.39	0.37	0.35	0.39
REGION				
Northeast	19.90	20.32	22.91	18.87
Southeast	24.07	22.89	21.87	24.23
Central	26.08	26.85	25.47	26.87
West	29.95	29.95	29.75	30.03
PARENTS' EDUCATION				
Less than high school	7.59	7.15	5.15	8.36
High school	21.32	19.69	18.44	21.47
Greater than high school	24.67	25.27	26.23	24.39
Graduated college	42.63	44.70	47.63	41.85
Unknown	3.24	2.67	2.12	3.35
TYPE OF LOCATION				
Large city	10.28	9.87	9.82	10.20
Midsize city	21.16	21.02	20.87	21.19
Urban fringe of large city	25.95	27.02	28.39	25.61
Urban fringe of midsize city	14.38	13.84	14.30	14.05
Large town	1.94	1.86	1.97	1.88
Small town	13.05	12.94	11.90	13.48
Rural/MSA	4.78	4.73	4.69	4.79
Rural/NonMSA	8.46	8.71	8.06	8.81
ESTIMATED TOTAL POPULATION	2841610	2545897	1638950	3748557

*****Note:*** *NAEP reporting subgroups are described in Appendix B.*

Table 19-52

Weighted Percentage of Students in the Geography Main Sample by Type of Eligibility and Subgroup Classification, Age 17/Grade 12*

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
SEX				
Male	50.77	50.39	46.11	52.55
Female	49.23	49.61	53.89	47.45
RACE/ETHNICITY				
White	70.30	74.45	76.24	70.51
Black	14.99	12.09	11.41	14.59
Hispanic	9.88	8.48	7.33	10.05
Asian American	2.27	2.60	2.42	2.43
Pacific Islander	1.09	1.19	1.38	1.03
American Indian	1.06	0.86	0.94	0.98
Unclassified	0.41	0.33	0.28	0.41
REGION				
Northeast	19.95	21.38	23.16	19.51
Southeast	24.21	22.53	21.83	24.11
Central	26.69	27.56	26.45	27.38
West	29.15	28.54	28.55	29.00
PARENTS' EDUCATION				
Less than high school	8.33	6.92	5.84	8.46
High school	21.07	21.44	19.31	22.09
Greater than high school	24.93	25.10	26.07	24.54
Graduated college	41.89	43.45	46.33	41.00
Unknown	3.25	2.53	1.96	3.33
TYPE OF LOCATION				
Large city	10.27	9.88	9.90	10.17
Midsize city	21.80	21.51	22.11	21.47
Urban fringe of large city	25.97	26.54	28.39	25.30
Urban fringe of midsize city	14.17	13.88	13.92	14.08
Large town	1.80	1.98	1.78	1.93
Small town	13.31	12.65	11.55	13.63
Rural/MSA	4.95	4.89	4.58	5.07
Rural/NonMSA	7.73	8.66	7.77	8.35
ESTIMATED TOTAL POPULATION	2840861	2542314	1641128	3742046

****Note***: *NAEP reporting subgroups are described in Appendix B.*

Table 19-53

Weighted Percentage of Students in the Reading and Writing Long-Term Trend Sample by Type of Eligibility and Subgroup Classification, Age 9/Grade 4*

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
SEX				
Male	49.81	50.26	46.85	51.40
Female	50.19	49.74	53.15	48.60
RACE/ETHNICITY				
White	70.21	68.52	70.17	69.03
Black	14.22	15.30	14.15	15.03
Hispanic	10.85	11.64	10.53	11.55
Asian American	2.03	2.14	2.61	1.86
American Indian	2.60	2.30	2.52	2.41
Unclassified	0.08	0.09	0.02	0.12
REGION				
Northeast	22.41	21.31	25.87	20.14
Southeast	24.33	24.42	22.58	25.14
Central	25.51	25.05	22.80	26.34
West	27.75	29.23	28.74	28.38
PARENTS' EDUCATION				
Less than high school	4.12	4.35	3.32	4.63
High school	15.62	16.04	15.14	16.12
Greater than high school	5.02	5.55	5.23	5.31
Graduated college	41.10	42.82	45.04	40.64
Unknown	33.81	30.90	30.94	32.97
TYPE OF LOCATION				
Large city	13.40	14.48	15.46	13.29
Midsize city	18.71	17.15	16.61	18.50
Urban fringe of large city	26.31	24.88	28.19	24.49
Urban fringe of midsize city	17.01	18.36	17.92	17.58
Large town	0.00	0.00	0.00	0.00
Small town	12.95	13.51	11.45	13.99
Rural/MSA	1.12	1.15	0.81	1.28
Rural/NonMSA	10.50	10.45	9.56	10.87
ESTIMATED TOTAL POPULATION	3494201	3479032	2088005	4885228

****Note:*** *NAEP reporting subgroups are described in Appendix B.*

Table 19-54

Weighted Percentage of Students in the Reading and Writing Long-Term Trend Sample by Type of Eligibility and Subgroup Classification, Age 13/Grade 8*

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
SEX				
Male	51.47	52.15	48.27	53.20
Female	48.53	47.85	51.73	46.80
RACE/ETHNICITY				
White	69.44	69.45	72.42	68.27
Black	14.24	15.25	11.43	16.05
Hispanic	11.41	10.45	10.91	10.94
Asian American	2.92	3.11	3.56	2.80
American Indian	1.83	1.62	1.48	1.82
Unclassified	0.17	0.12	0.21	0.12
REGION				
Northeast	20.40	20.70	24.85	18.86
Southeast	21.77	23.27	20.30	23.39
Central	26.79	25.71	23.88	27.19
West	31.03	30.31	30.96	30.56
PARENTS' EDUCATION				
Less than high school	6.93	6.93	5.48	7.51
High school	26.89	26.90	24.22	27.95
Greater than high school	10.52	11.83	12.11	10.80
Graduated college	46.31	45.95	50.80	44.30
Unknown	8.61	7.89	6.96	8.76
TYPE OF LOCATION				
Large city	13.38	13.74	12.89	13.82
Midsize city	19.39	18.79	17.28	19.80
Urban fringe of large city	21.14	20.92	23.50	20.05
Urban fringe of midsize city	20.62	20.83	21.80	20.30
Large town	0.00	0.00	0.00	0.00
Small town	12.85	13.49	11.87	13.68
Rural/MSA	2.81	2.69	2.83	2.72
Rural/NonMSA	9.81	9.56	9.82	9.63
ESTIMATED TOTAL POPULATION	3437249	3401434	1933571	4905112

*Note: *NAEP reporting subgroups are described in Appendix B.*

Table 19-55

Weighted Percentage of Students in the Reading and Writing Long-Term Trend Sample by Type of Eligibility and Subgroup Classification, Age 17/Grade 11*

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
SEX				
Male	50.23	48.46	47.03	50.43
Female	49.77	51.54	52.97	49.57
RACE/ETHNICITY				
White	70.17	68.32	75.51	66.34
Black	14.99	14.70	11.40	16.45
Hispanic	10.19	11.90	8.67	12.14
Asian American	3.15	3.63	3.01	3.57
American Indian	1.38	1.23	1.21	1.35
Unclassified	0.13	0.22	0.20	0.16
REGION				
Northeast	22.07	22.22	21.14	22.61
Southeast	24.83	22.93	23.36	24.13
Central	26.02	26.22	27.37	25.53
West	27.08	28.63	28.12	27.72
PARENTS' EDUCATION				
Less than high school	7.31	7.51	5.42	8.34
High school	27.22	25.88	24.77	27.39
Greater than high school	19.11	19.71	20.10	19.08
Graduated college	42.77	43.62	47.34	41.27
Unknown	3.31	3.03	2.07	3.68
TYPE OF LOCATION				
Large city	14.91	16.69	13.10	17.05
Midsize city	16.14	14.67	15.45	15.39
Urban fringe of large city	21.17	20.52	20.78	20.87
Urban fringe of midsize city	23.17	23.12	24.64	22.45
Large town	0.00	0.00	0.00	0.00
Small town	13.51	13.99	14.08	13.60
Rural/MSA	1.70	1.78	1.88	1.67
Rural/NonMSA	9.40	9.23	10.07	8.96
ESTIMATED TOTAL POPULATION	3124711	3104689	1977526	4251875

****Note:*** *NAEP reporting subgroups are described in Appendix B.*

Table 19-56

Weighted Percentage of Students in the Mathematics and Science Long-Term Trend Sample by Type of Eligibility and Subgroup Classification, Age 9/Grade 4*

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
SEX				
Male	48.89	46.85	46.85	48.89
Female	51.11	53.15	53.15	51.11
RACE/ETHNICITY				
White	70.11	69.67	69.67	70.11
Black	14.35	15.66	15.66	14.35
Hispanic	10.95	9.99	9.99	10.95
Asian American	2.07	2.48	2.48	2.07
American Indian	2.33	2.00	2.00	2.33
Unclassified	0.21	0.19	0.19	0.21
REGION				
Northeast	21.77	24.47	24.47	21.77
Southeast	24.55	24.88	24.88	24.55
Central	25.93	23.94	23.94	25.93
West	27.75	26.71	26.71	27.75
PARENTS' EDUCATION				
Less than high school	3.83	3.14	3.14	3.83
High school	13.97	13.75	13.75	13.97
Greater than high school	7.00	7.66	7.66	7.00
Graduated college	44.91	47.43	47.43	44.91
Unknown	30.12	27.92	27.92	30.12
TYPE OF LOCATION				
Large city	14.32	16.73	16.73	14.32
Midsize city	17.58	16.63	16.63	17.58
Urban fringe of large city	26.58	26.95	26.95	26.58
Urban fringe of midsize city	17.48	18.34	18.34	17.48
Large town	0.00	0.00	0.00	0.00
Small town	13.16	11.74	11.74	13.16
Rural/MSA	0.93	0.84	0.84	0.93
Rural/NonMSA	9.94	8.77	8.77	9.94
ESTIMATED TOTAL POPULATION	3308719	2196784	2196784	3308719

****Note***: *NAEP reporting subgroups are described in Appendix B.*

Table 19-57

Weighted Percentage of Students in the Mathematics and Science Long-Term Trend Sample by Type of Eligibility and Subgroup Classification, Age 13/Grade 8*

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
SEX				
Male	49.20	45.31	45.31	49.20
Female	50.80	54.69	54.69	50.80
RACE/ETHNICITY				
White	69.74	72.65	72.65	69.74
Black	14.75	12.86	12.86	14.75
Hispanic	10.77	9.97	9.97	10.77
Asian American	3.11	3.38	3.38	3.11
American Indian	1.28	0.77	0.77	1.28
Unclassified	0.34	0.37	0.37	0.34
REGION				
Northeast	20.17	22.46	22.46	20.17
Southeast	22.21	21.55	21.55	22.21
Central	26.59	25.62	25.62	26.59
West	31.04	30.37	30.37	31.04
PARENTS' EDUCATION				
Less than high school	6.17	5.20	5.20	6.17
High school	23.00	21.95	21.95	23.00
Greater than high school	16.74	17.64	17.64	16.74
Graduated college	45.25	48.32	48.32	45.25
Unknown	8.28	6.50	6.50	8.28
TYPE OF LOCATION				
Large city	12.89	13.78	13.78	12.89
Midsize city	19.01	17.29	17.29	19.01
Urban fringe of large city	21.85	23.91	23.91	21.85
Urban fringe of midsize city	19.97	20.68	20.68	19.97
Large town	0.00	0.00	0.00	0.00
Small town	13.02	12.84	12.84	13.02
Rural/MSA	2.71	2.71	2.71	2.71
Rural/NonMSA	10.54	8.78	8.78	10.54
ESTIMATED TOTAL POPULATION	3353585	2063833	2063833	3353585

**Note: NAEP reporting subgroups are described in Appendix B.*

Table 19-58

Weighted Percentage of Students in the Mathematics and Science Long-Term Trend Sample by Type of Eligibility and Subgroup Classification, Age 17/Grade 11*

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
SEX				
Male	49.48	47.23	47.23	49.48
Female	50.52	52.77	52.77	50.52
RACE/ETHNICITY				
White	70.53	73.79	73.79	70.53
Black	14.79	12.80	12.80	14.79
Hispanic	10.10	9.12	9.12	10.10
Asian American	3.05	3.01	3.01	3.05
American Indian	1.11	0.91	0.91	1.11
Unclassified	0.42	0.36	0.36	0.42
REGION				
Northeast	22.75	22.20	22.20	22.75
Southeast	24.61	24.48	24.48	24.61
Central	25.55	25.58	25.58	25.55
West	27.09	27.74	27.74	27.09
PARENTS' EDUCATION				
Less than high school	6.94	5.84	5.84	6.94
High school	22.01	19.94	19.94	22.01
Greater than high school	23.47	25.17	25.17	23.47
Graduated college	44.24	46.49	46.49	44.24
Unknown	3.06	2.40	2.40	3.06
TYPE OF LOCATION				
Large city	15.65	14.22	14.22	15.65
Midsize city	13.97	14.00	14.00	13.97
Urban fringe of large city	19.84	19.95	19.95	19.84
Urban fringe of midsize city	23.90	24.24	24.24	23.90
Large town	0.00	0.00	0.00	0.00
Small town	14.71	15.37	15.37	14.71
Rural/MSA	1.76	1.76	1.76	1.76
Rural/NonMSA	10.18	10.47	10.47	10.18
ESTIMATED TOTAL POPULATION	3078721	2249986	2249986	3078721

****Note***: *NAEP reporting subgroups are described in Appendix B.*

Table 19-59

Weighted Percentage of Excluded Students in the Main Samples
by Type of Eligibility and Subgroup Classification
Age 9/Grade 4

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
SEX				
Male	63.05	66.82	62.61	65.42
Female	36.95	33.18	37.39	34.58
RACE/ETHNICITY				
White	52.08	49.22	47.44	51.51
Black	23.97	24.04	20.23	24.86
Hispanic	19.28	23.26	27.08	19.79
Other	4.54	3.43	5.11	3.76
REGION				
Northeast	22.08	21.28	24.07	21.16
Southeast	21.29	24.12	18.95	23.51
Central	22.91	21.34	19.49	22.74
West	33.72	33.26	37.50	32.59
ESTIMATED TOTAL POPULATION	215287	204621	80652	339858

Table 19-60

Weighted Percentage of Excluded Students in the Main Samples
by Type of Eligibility and Subgroup Classification
Age 13/Grade 8

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
SEX				
Male	68.70	68.35	65.93	68.83
Female	31.30	31.65	34.07	31.17
RACE/ETHNICITY				
White	54.96	55.83	57.86	55.00
Black	24.96	21.70	14.96	24.57
Hispanic	16.86	17.70	23.24	16.38
Other	3.14	4.60	3.66	3.94
REGION				
Northeast	22.48	21.05	26.54	21.11
Southeast	20.04	22.04	13.54	22.07
Central	27.67	27.67	18.95	28.98
West	29.81	29.24	40.97	27.85
ESTIMATED TOTAL POPULATION	163235	166808	41479	289332

Table 19-61

Weighted Percentage of Excluded Students in the Main Samples by Type of Eligibility and Subgroup Classification

Age 17/Grade 12

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
SEX				
Male	66.30	61.80	62.53	64.62
Female	33.70	38.20	37.47	35.38
RACE/ETHNICITY				
White	54.75	67.57	68.99	59.07
Black	20.90	15.90	9.82	20.01
Hispanic	19.15	11.17	15.82	15.68
Other	4.96	5.19	5.25	5.03
REGION				
Northeast	24.56	22.87	33.33	22.57
Southeast	21.55	23.38	15.15	23.39
Central	23.94	26.27	23.81	25.06
West	29.95	27.47	27.71	28.98
ESTIMATED TOTAL POPULATION	112840	86163	23113	176182

Table 19-62

Weighted Percentage of Excluded Students in the Long-Term Trend Samples by Type of Eligibility and Subgroup Classification

Age 9/Grade 4

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
SEX				
Male	62.19	64.59	59.91	64.07
Female	37.81	35.41	40.09	35.93
RACE/ETHNICITY				
White	51.94	55.51	43.97	55.68
Black	21.12	18.55	14.89	20.91
Hispanic	18.19	19.65	29.57	16.68
Other	8.75	6.29	11.57	6.73
REGION				
Northeast	17.28	17.57	22.85	16.29
Southeast	27.61	24.58	15.66	28.32
Central	18.49	16.82	15.80	18.07
West	36.62	41.03	45.70	37.31
ESTIMATED TOTAL POPULATION	278296	260863	92699	446460

Table 19-63

Weighted Percentage of Excluded Students in the Long-Term Trend Samples by Type of Eligibility and Subgroup Classification

Age 13/Grade 8

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
SEX				
Male	63.02	66.12	60.14	65.19
Female	36.98	33.88	39.86	34.81
RACE/ETHNICITY				
White	51.42	52.47	45.31	52.91
Black	21.30	20.50	13.70	21.95
Hispanic	21.36	21.71	29.25	20.41
Other	5.92	5.32	11.75	4.73
REGION				
Northeast	15.34	10.81	14.74	12.85
Southeast	23.02	22.25	16.19	23.58
Central	27.26	31.16	18.95	30.68
West	34.39	35.78	50.12	32.89
ESTIMATED TOTAL POPULATION	221212	217144	55696	382661

Table 19-64

Weighted Percentage of Excluded Students in the Long-Term Trend Samples by Type of Eligibility and Subgroup Classification

Age 17/Grade 11

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
SEX				
Male	63.97	64.44	58.66	65.18
Female	36.03	35.56	41.34	34.82
RACE/ETHNICITY				
White	59.06	53.89	63.12	55.42
Black	17.86	22.04	14.22	20.86
Hispanic	18.04	19.89	17.71	19.14
Other	5.04	4.18	4.95	4.57
REGION				
Northeast	26.25	21.30	21.45	24.31
Southeast	23.96	26.80	21.07	26.08
Central	21.96	17.94	20.72	19.91
West	27.83	33.96	36.77	29.70
ESTIMATED TOTAL POPULATION	171136	156898	49570	278464

Table 19-65

Number of Students in the Main Samples with Proficiency Scores, by Type of Eligibility

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
AGE 9/GRADE 4				
Reading	7042	7382	4744	9680
U.S. History	5239	5499	3533	7205
Geography	5167	5507	3469	7205
Total	17448	18388	11746	24090
AGE 13/GRADE 8				
Reading	9683	10135	6271	13547
U.S. History	8302	8767	5400	11669
Geography	6558	6878	4250	9186
Total	24543	25780	15921	34402
AGE 17/GRADE 12				
Reading	10301	9935	6531	13705
U.S. History	8132	7818	5113	10837
Geography	6473	6234	4118	8589
Total	24906	23987	15762	33131

Table 19-66

Number of Students in the Long-Term Trend Samples with Proficiency Scores, by Type of Eligibility

	Eligible by			
	Age	Grade	Age & Grade	Age or Grade
AGE 9/GRADE 4				
Reading	3871	4098	2634	5335
Writing	3205	3409	2190	4424
Mathematics	5663	3751	3751	5663
Science	5663	3751	3751	5663
Total	9534	7849	6385	10998
AGE 13/GRADE 8				
Reading	3989	4058	2500	5547
Writing	3989	4058	2500	5547
Mathematics	6052	3824	3824	6052
Science	6052	3824	3824	6052
Total	10041	7882	6324	11599
AGE 17/GRADE 12				
Reading	3811	3948	2919	4840
Writing	3811	3948	2919	4840
Mathematics	3813	2832	2832	3813
Science	3813	2832	2832	3813
Total	7624	6780	5751	8653

Table 19-67

Raw Counts and Total Unweighted Numbers of IEP and LEP Students
Age 9/Grade 4

	Age	Grade	Age & Grade	Total
ASSESSED STUDENTS ONLY				
IEP	253	306	143	416
LEP	186	199	131	254
Both IEP and LEP	7	7	5	10
Total Sample	17448	18388	11746	24090
EXCLUDED STUDENTS ONLY				
IEP	608	685	264	1031
LEP	364	368	216	517
Both IEP and LEP	17	28	9	36
Total Sample	955	1025	471	1512
ASSESSED AND EXCLUDED COMBINED				
IEP	861	991	407	1447
LEP	550	567	347	771
Both IEP and LEP	24	35	14	46
Total Sample	18403	19413	12217	25602
IEP STUDENTS				
Excluded	608	685	264	1031
Assessed	253	306	143	416
Total Sample	18403	19413	12217	25602
LEP STUDENTS				
Excluded	364	368	216	517
Assessed	186	199	131	254
Total Sample	18403	19413	12217	25602
IEP OR LEP STUDENTS				
Excluded	955	1025	471	1512
Assessed	432	498	270	660
Total Sample	18403	19413	12217	25602
IEP AND LEP STUDENTS				
Excluded	17	28	9	36
Assessed	7	7	5	10
Total Sample	18403	19413	12217	25602

Table 19-68

Raw Counts and Total Unweighted Numbers of IEP and LEP Students
Age 13/Grade 8

	Age	Grade	Age & Grade	Total
ASSESSED STUDENTS ONLY				
IEP	431	448	145	734
LEP	178	189	103	264
Both IEP and LEP	9	7	4	12
Total Sample	29691	31251	19214	41728
EXCLUDED STUDENTS ONLY				
IEP	893	979	253	1621
LEP	253	323	112	467
Both IEP and LEP	27	24	7	44
Total Sample	1119	1278	358	2044
ASSESSED AND EXCLUDED COMBINED				
IEP	1324	1427	398	2355
LEP	431	512	215	731
Both IEP and LEP	36	31	11	56
Total Sample	30810	32529	19572	43772
IEP STUDENTS				
Excluded	893	979	253	1621
Assessed	431	448	145	734
Total Sample	30810	32529	19572	43772
LEP STUDENTS				
Excluded	253	323	112	467
Assessed	178	189	103	264
Total Sample	30810	32529	19572	43772
IEP OR LEP STUDENTS				
Excluded	1119	1278	358	2044
Assessed	600	630	245	986
Total Sample	30810	32529	19572	43772
IEP AND LEP STUDENTS				
Excluded	27	24	7	44
Assessed	9	7	4	12
Total Sample	30810	32529	19572	43772

Table 19-69

Raw Counts and Total Unweighted Numbers of IEP and LEP Students
Age 17/Grade 12

	Age	Grade	Age & Grade	Total
ASSESSED STUDENTS ONLY				
IEP	311	284	106	490
LEP	192	172	71	294
Both IEP and LEP	6	3	2	7
Total Sample	29273	28233	18560	38946
EXCLUDED STUDENTS ONLY				
IEP	889	776	193	1474
LEP	316	184	76	425
Both IEP and LEP	15	12	2	25
Total Sample	1190	948	267	1874
ASSESSED AND EXCLUDED COMBINED				
IEP	1200	1060	299	1964
LEP	508	356	147	719
Both IEP and LEP	21	15	4	32
Total Sample	30463	29181	18827	40820
IEP STUDENTS				
Excluded	889	776	193	1474
Assessed	311	284	106	490
Total Sample	30463	29181	18827	40820
LEP STUDENTS				
Excluded	316	184	76	425
Assessed	192	172	71	294
Total Sample	30463	29181	18827	40820
IEP OR LEP STUDENTS				
Excluded	1190	948	267	1874
Assessed	498	453	174	776
Total Sample	30463	29181	18827	40820
IEP AND LEP STUDENTS				
Excluded	15	12	2	25
Assessed	6	3	2	7
Total Sample	30463	29181	18827	40820

Table 19-70

Weighted Percentages and Total Weighted Numbers of IEP and LEP Students
Age 9/Grade 4

	Age	Grade	Age & Grade	Total
ASSESSED STUDENTS ONLY				
IEP	4.66	5.81	3.97	5.76
LEP	2.37	2.49	2.44	2.42
Both IEP and LEP	0.08	0.08	0.07	0.09
Total Sample	3555583.26	3528673.33	2092000.08	4992256.51
EXCLUDED STUDENTS ONLY				
IEP	76.67	76.82	68.43	78.72
LEP	24.69	25.53	33.32	23.14
Both IEP and LEP	1.36	2.36	1.75	1.86
Total Sample	215287.33	204621.02	80652.22	339858.46
ASSESSED AND EXCLUDED COMBINED				
IEP	8.77	9.70	6.36	10.41
LEP	3.64	3.75	3.59	3.74
Both IEP and LEP	0.16	0.20	0.13	0.20
Total Sample	3770870.59	3733294.35	2172652.30	5332114.97
IEP STUDENTS				
Excluded	49.90	43.39	39.91	48.19
Assessed	50.10	56.61	60.09	51.81
Total Sample	330749.01	362278.17	138287.20	555217.75
LEP STUDENTS				
Excluded	38.71	37.33	34.48	39.43
Assessed	61.29	62.67	65.52	60.57
Total Sample	137300.80	139978.98	77936.80	199467.54
IEP OR LEP STUDENTS				
Excluded	46.58	41.37	37.80	45.68
Assessed	53.42	58.63	62.20	54.32
Total Sample	462185.90	494606.71	213387.12	744007.82
IEP AND LEP STUDENTS				
Excluded	49.76	63.11	49.78	59.32
Assessed	50.24	36.89	50.22	40.68
Total Sample	5863.90	7650.44	2836.88	10677.46

Table 19-71

Weighted Percentages and Total Weighted Numbers of IEP and LEP Students
Age 13/Grade 8

	Age	Grade	Age & Grade	Total
ASSESSED STUDENTS ONLY				
IEP	4.94	4.67	2.19	5.84
LEP	1.49	1.48	1.36	1.53
Both IEP and LEP	0.07	0.04	0.04	0.06
Total Sample	3526670.97	3448320.50	1974386.80	5000604.67
EXCLUDED STUDENTS ONLY				
IEP	86.20	82.70	77.87	85.28
LEP	15.94	18.94	23.53	16.68
Both IEP and LEP	2.14	1.64	1.40	1.95
Total Sample	163234.85	166807.54	41479.02	289332.40
ASSESSED AND EXCLUDED COMBINED				
IEP	8.53	8.27	3.74	10.19
LEP	2.13	2.28	1.82	2.36
Both IEP and LEP	0.16	0.12	0.07	0.17
Total Sample	3689905.82	3615128.04	2015865.82	5289937.07
IEP STUDENTS				
Excluded	44.70	46.12	42.80	45.79
Assessed	55.30	53.88	57.20	54.21
Total Sample	314793.58	299124.94	75461.48	538821.20
LEP STUDENTS				
Excluded	33.18	38.30	26.67	38.69
Assessed	66.82	61.70	73.33	61.31
Total Sample	78416.83	82466.92	36594.08	124694.53
IEP OR LEP STUDENTS				
Excluded	42.15	44.20	37.50	44.19
Assessed	57.85	55.80	62.50	55.81
Total Sample	387231.64	377396.19	110624.20	654772.66
IEP AND LEP STUDENTS				
Excluded	58.46	65.24	40.50	64.65
Assessed	41.54	34.76	59.50	35.35
Total Sample	5978.77	4195.67	1431.37	8743.07

Table 19-72

Weighted Percentages and Total Weighted Numbers of IEP and LEP Students
Age 17/Grade 12

	Age	Grade	Age & Grade	Total
ASSESSED STUDENTS ONLY				
IEP	3.73	3.41	1.91	4.31
LEP	1.68	1.33	0.77	1.84
Both IEP and LEP	0.05	0.02	0.02	0.05
Total Sample	2846149.13	2559298.68	1645024.11	3760423.70
EXCLUDED STUDENTS ONLY				
IEP	81.36	87.16	80.05	84.37
LEP	19.92	13.66	20.36	16.80
Both IEP and LEP	1.28	0.83	0.41	1.17
Total Sample	112839.90	86163.39	23112.80	176182.41
ASSESSED AND EXCLUDED COMBINED				
IEP	6.69	6.13	2.99	7.89
LEP	2.38	1.73	1.04	2.51
Both IEP and LEP	0.10	0.05	0.03	0.10
Total Sample	2958989.03	2645462.07	1668136.91	3936606.11
IEP STUDENTS				
Excluded	46.36	46.28	37.07	47.85
Assessed	53.64	53.72	62.93	52.15
Total Sample	198028.56	162282.76	49917.61	310633.54
LEP STUDENTS				
Excluded	31.95	25.77	27.17	29.97
Assessed	68.05	74.23	72.83	70.03
Total Sample	70347.36	45689.78	17319.26	98769.98
IEP OR LEP STUDENTS				
Excluded	42.51	41.70	34.61	43.44
Assessed	57.49	58.30	65.39	56.56
Total Sample	265416.85	206646.02	66784.05	405570.74
IEP AND LEP STUDENTS				
Excluded	48.90	53.84	21.15	53.89
Assessed	51.10	46.16	78.85	46.11
Total Sample	2959.08	1326.52	452.82	3832.78

Appendix A

PARTICIPANTS IN THE OBJECTIVES AND ITEM DEVELOPMENT PROCESS

Appendix A

PARTICIPANTS IN THE OBJECTIVES AND ITEM DEVELOPMENT PROCESS

READING PROJECT STEERING COMMITTEE

Gary Marx	American Association of School Administrators, Arlington, Virginia
Carole Perlman	American Educational Research Association, Chicago, Illinois
Marilyn Rauth	American Federation of Teachers, Washington, D.C.
Edward Roeber	Association of State Assessment Programs, Lansing, Michigan
Helene Hodges	Association of Supervision and Curriculum Development, Alexandria, Virginia
H. Dean Evans	Council of Chief State School Officers, Indianapolis, Indiana
Esther Schaeffer	National Alliance of Business, Washington, D.C.
Kathleen Holliday	National Association of Elementary School Principals, Potomac, Maryland
Ann Smith	National Educational Association, Ormond Beach, Florida
Mike Cohen	National Governors' Association, Washington, D.C.
Ann Kahn	National Parent Teacher Association, Alexandria, Virginia
Scott Thompson	National Education of Secondary School Principals, Reston, Virginia
Harriet C. Jelnek	National School Board Association, Rhineland, Wisconsin
Paul Le Mahieu	National Association of Test Directors, Pittsburgh, Pennsylvania
Br. Robert Kealey	National Catholic Educational Association, Washington, D.C.

READING PROJECT PLANNING COMMITTEE

Marilyn Adams	BBN Laboratories, Cambridge, Massachusetts
Marsha Delain	South Carolina Department of Education, Columbia, South Carolina
Lisa Delpit	Institute for Urban Research, Baltimore, Maryland

BEST COPY AVAILABLE

READING PROJECT PLANNING COMMITTEE, continued

William Feehan	Chase Manhattan Bank, New York, New York
Philip Gough	University of Texas at Austin, Austin, Texas
Edward Haertel	Stanford University, Stanford, California
Elfrieda Hiebert	University of Colorado, Boulder, Colorado
Judith Langer	State University of New York, Albany, New York
P. David Pearson	University of Illinois, Champaign, Illinois
Charles Peters	Oakland Schools, Pontiac, Michigan
John P. Pikulski	University of Delaware, Newark, Delaware
Keith Stanovich	Oakland University, Rochester, Michigan
Paul Randy Walker	Maine Department of Education, Augusta, Maine
Sheila Valencia	University of Washington, Seattle, Washington
Janet Jones	Charles County Public Schools, Waldorf, Maryland

1992 NAEP READING CCSSO PROJECT STAFF

Ramsay W. Selden	Director, State Education Assessment Center, Council of Chief State School Officers
Barbara Kapinus	Project Coordinator
Diane Schilder	Project Associate

1994 READING ITEM DEVELOPMENT COMMITTEE

Dr. Katherine H. Au	Honolulu, Hawaii
Carmela Cocola	Yardley, Pennsylvania
Dr. Janice Dole	University of Utah, Salt Lake City, Utah
Dr. Alan Farstrup	International Reading Association, Newark, Delaware

1994 READING ITEM DEVELOPMENT COMMITTEE, continued

Herberto Godina	Champaign, Illinois
Dr. Susan Hynds	Syracuse University, Syracuse, New York
Dr. Barbara Kapinus	Council of Chief State School Officers, Washington, D.C.
Dr. Judith Langer	State University of New York, Albany, New York
Dr. Susan Neuman	Temple University, Philadelphia, Pennsylvania
Dr. David Pearson	University of Illinois, Champaign, Illinois
Dr. Jesse Perry	San Diego, California
Dr. John Pikulski	University of Delaware, Newark, Delaware
Dr. Timothy Shanahan	University of Illinois at Chicago, Chicago, Illinois
Laura Tsosie	Pinon, Arizona

U.S. HISTORY PROJECT STEERING COMMITTEE MEMBERS

Margorie Bingham	St. Louis Park Senior High School, Minnetonka, Minnesota
Austin Creel	University of Florida, Department of Religion, Gainesville, Florida
Thomas S. Dickenson	*Middle School Journal*, National Middle School Association, Columbus, Ohio
Tom Dunthorn	Florida Department of Education, Tallahassee, Florida
Gerald M. Eads II	Department of Assessment and Testing, Virginia State Department of Education, Richmond, Virginia
William Everdell	Saint Ann's School, Brooklyn Heights, New York
Elizabeth Fox-Genovese	Institute of Women's Studies, Emory University, Atlanta, Georgia
Sandra French	National PTA, Ellicott City, Maryland
Art Gosling	Arlington County Schools, Arlington, Virginia
Byron Hollinshead	*Americana Magazine, Inc.*, New York, New York

U.S. HISTORY PROJECT STEERING COMMITTEE MEMBERS, continued

Bill Honig	California State Department of Education, Sacramento, California
Arnita A. Jones	Organization of American Historians, Bloomington, Indiana
Leon Litwack	University of California, Berkely, California
Ruth Wattenberg	Education for Democracy Project, American Federation of Teachers, Washington, D.C.
Robert D. Reynolds, Jr.	The George Meany Memorial Archives, Silver Spring, Maryland
Armstead L. Robinson	University of Virginia, Charlottesville, Virginia
Everett V. Samuelson	College of Education, University of Idaho, Moscow, Idaho
Diane Stallings	Colonial National Historical Park, Newport News, Virginia
August Steinhilber	National School Boards Association, Alexandria, Virginia
Kathleen Hunter	National Trust for Historic Preservation, Washington, D.C.
Caryolyn Yoder	*Cobblestone Magazine*, Peterborough, New Hampshire
Sandra Harp	Minneapolis Board of Education, Minneapolis, Minnesota
Karol Wolgemuth	Greenview Knolls Elementary School, Great Mills, Maryland
James B. Gardner-Ex Officio	American Historical Association, Washington, D.C.
Frances Haley-Ex Officio	National Council for the Social Studies, Washington, D.C.
Elaine Reed*-Ex Officio	National Council for History Education, Westlake, Ohio

*(*replaced Paul Gagnon on 12/2/91)*

U.S. HISTORY PROJECT PLANNING COMMITTEE MEMBERS

Carol Berkin	Baruch College, Department of History, New York, New York
William Branch	Evanston Township High School #202, Bellwood, Illinois
Stewart Caffey	Jefferies Junior High, Comanche Independent School District, Comanche, Texas

U.S. HISTORY PROJECT PLANNING COMMITTEE MEMBERS, continued

Pedro Castillo	University of California, Oakes College, Santa Cruz, California
Charlotte Crabtree*	National Center for History in the Schools, University of California, Los Angeles, California
Alberta Sebolt George	Sturbridge Village, Sturbridge, Massachusetts
George M. Gregory	New York State Education Department, Albany, New York
Claudia Hoone	Elementary School #58, Indianapolis, Indiana
Jean Jamgochian	Haycock Gifted and Talented Center, Falls Church, Virginia
Richard Kirkendall	University of Washington, History Department, Seattle, Washington
Linda Levstik	University of Kentucky, Lexington, Kentucky
Tedd Levy	Nathan Hale Middle School, Norwalk, Connecticut
Earl Lewis	Center for Afro-American and African Studies, University of Michigan, Ann Arbor, Michigan
Glenabah Martinez	Rio Grande High School, Albuquerque, New Mexico
John Patrick	Indiana University, Bloomington, Indiana
Robert Summerville	Alabama State Department of Education, Secondary Instructional Services, Montgomery, Alabama
Joyce Stevos	Providence School Department, Providence, Rhode Island
Stephan Thernstrom	Harvard University, Cambridge, Massachusetts
Lawana Trout	Central State University, Chicago, Illinois
Reed Ueda	Tufts University, Medford, Massachusetts
Jon Wakelyn	Catholic University, Washington, D.C.
Ronald G. Walters	Johns Hopkins University, Baltimore, Maryland

*(*Served on Planning Committee from September 1991 - March 1992. Appointed to National Assessment Governing Board on 3/5/92)*

1994 U.S. HISTORY ITEM DEVELOPMENT COMMITTEE

Keith Barton	Winchester, Kentucky
Michael Beck	Beta, Inc., Pleasantville, New York
Carol Berkin	Baruch College, Brooklyn, New York
Liz Burton	University of Colorado, Boulder, Colorado
Sam Caruso	Tucson, Arizona
William Cirullo	Pennington, New Jersey
Charles Cook	St. John's School, Houston, Texas
John Crum	Wilmington, Delaware
Jane Dailey	Houston, Texas
Denise Driscoll	Stuart County Day School, Princeton, New Jersey
Alison Drew	North Brunswick, New Jersey
Anthony Edmonds	Ball State University, Muncie, Indiana
Barbara Hinkley	St. Charles West High School, St. Charles, Missouri
Claudia Hoone	Indianapolis, Indiana
Nancy Karweit	Baltimore, Maryland
Milton Kato	Thomas Jefferson Jr. High School, Madera, California
Karen Latham	Hopewell, New Jersey
Linda Levstik	University of Kentucky, Lexington, Kentucky
Tedd Levy	Nathan Hale Middle School, Norwalk, Connecticut
Glenabah Martinez	Rio Grande High School, Albuquerque, New Mexico
Anne Medio	Chicago, Illinois
Judith Mohraz	Southern Methodist University, Dallas, Texas
Lori Morton	Lynbrook School, Springfield, Virginia

1994 U.S. HISTORY ITEM DEVELOPMENT COMMITTEE, continued

John Patrick	Indiana University, Bloomington, Indiana
Eileen Randall	Trenton, New Jersey
Stephan Ternstrom	Harvard University, Cambridge, Massachusetts
Andrew Weiss	Princeton, New Jersey
Susan Wickley	Robbinsville, New Jersey
Alan Winkler	Miami University, Oxford, Ohio
Thomas Wissinger	Thomas Harrington Middle School, Mt. Laurel, New Jersey

GEOGRAPHY PROJECT STEERING COMMITTEE MEMBERS

Frank Betts	Association for Supervision and Curriculum Development, Alexandria, Virginia
Admiral Geoffrey Chesbrough	U.S. Naval Observatory, Washington, D.C.
Anthony R. de Souza	International Geographical Congress, Washington, D.C.
Allan Eustis	Channel 4, WRC-TV, Washington, D.C.
Victor H. Ferry	National Association of Elementary School Principals, Waterford, Connecticut
Gilbert M. Grosvenor	National Geographic Society, Washington, D.C.
The Honorable Samuel E. Hayes, Jr.	Pennsylvania House of Representatives, Harrisburg, Pennsylvania
Joan Herman	Center for Research on Evaluation, Standards, and Student Testing, University of California, Los Angeles, California
A. David Hill	University of Colorado, Boulder, Colorado
Tom Koerner	National Association of Secondary School Principals, Reston, Virginia
Sister Catherine McNamee	National Catholic Educational Association, Washington, D.C.
Paul M. Ostergard	Citibank, New York, New York

GEOGRAPHY PROJECT STEERING COMMITTEE MEMBERS, continued

Sharon Purcell	National Council for the Social Studies, Washington, D.C.
Mark Root	Council of Great City Schools, Washington, D.C.
Harold P. Seamon	National School Boards Association, Alexandria, Virginia
Brock Stanton	IBM, Atlanta, Georgia
Walt Tremer	National Education Association, Coopersburg, Pennsylvania
Manya S. Ungar	National PTA, Scotch Plains, New Jersey
Ruth Wattenberg	American Federation of Teachers, Washington, D.C.
Gene Wilhoit	National Association of State Boards of Education, Alexandria, Virginia

GEOGRAPHY PROJECT PLANNING COMMITTEE MEMBERS

Sarah Bednarz	Journal of Geography, College Station, Texas
Norman C. Bettis	National Council for Geographic Education, Normal, Illinois
Richard Boehm	Southwest Texas State University, San Marcos, Texas
Roger Downs	Pennsylvania State University, University Park, Pennsylvania
Charles Fitzpatrick	St. Paul Academy & Summit School, St. Paul, Minnesota
Frances Gatz	International Expeditions, Alexandria, Virginia
Susan Hardwick	National Council for Geographic Education, Chico, California
Carter B. Hart, Jr.	Concord, New Hampshire
Rita Duarte Herrera	Alum Rock Union School District, San Jose, California
Gail Ludwig	Missouri Geographic Alliance Coordinator, Columbia, Missouri
Melvin G. Marcus	Arizona State University, Tempe, Arizona
James Marran	New Trier High School, Winnetka, Illinois
Robert Morrill	Virginia Tech, Blacksburg, Virginia
Barbara Moses	School District of Philadelphia, Philadelphia, Pennsylvania

GEOGRAPHY PROJECT PLANNING COMMITTEE MEMBERS, continued

Christopher L. Salter	University of Missouri, Columbia, Missouri
Sharif Shakrani	Michigan Department of Education, Lansing, Michigan
William Strong	National Geographic Society, Washington, D.C.
Thomas J. Wilbanks	Developing Country Programs, Oak Ridge National Laboratory, Oak Ridge, Tennessee
Barbara Winston	Northeastern Illinois University, Chicago, Illinois

1994 GEOGRAPHY ITEM DEVELOPMENT COMMITTEE

Michael Beck	Beta, Inc., Pleasantville, New York
Norman Bettis	Illinois State University, Normal, Illinois
Stan Christodlous	Ledgewood, New Jersey
Alison Drew	North Brunswick, New Jersey
Susan Hanson	Clark University, Worcester, Massachusetts
Rita Herrera	Ocala Middle School, San Jose, California
Briavel Holcomb	Rutgers University, New Brunswick, New Jersey
Kathleen Lynch	Glenside, Pennsylvania
Vincent Malmstrom	Dartmouth College, Hanover, New Hampshire
James Marran	New Trier High School, Winnetka, Illinois
Susan Monroe	National Council for Geographic Education, Washington, D.C.
Barbara Moses	School District of Philadelphia, Philadelphia, Pennsylvania
Salvatore Natoli	National Council for Social Studies, Washington, D.C.
Christopher Salter	University of Missouri, Columbia, Missouri
Harley Walker	Louisiana State University, Baton Rouge, Louisiana

BEST COPY AVAILABLE

1994 GEOGRAPHY ITEM DEVELOPMENT COMMITTEE, continued

Susan Wickley	Robbinsville, New Jersey
Barbara Winston	Northern Illinois University, Chicago, Illinois
Thomas Wissinger	Thomas Harrington Middle School, Mt. Laurel, New Jersey
Milbrey Zelley	Metuchen High School, Metuchen, New Jersey

Appendix B

REPORTING SUBGROUPS
COMPOSITE AND DERIVED COMMON BACKGROUND VARIABLES
COMPOSITE AND DERIVED REPORTING VARIABLES

Appendix B

REPORTING SUBGROUPS FOR THE 1994 NAEP ASSESSMENT

Results for the 1994 assessment were reported for student subgroups defined by gender, race/ethnicity, type of location, parents' level of education, and geographical region. The following explains how each of these subgroups was derived.

DSEX (Gender)

The variable SEX is the gender of the student being assessed, as taken from school records. For a few students, data for this variable was missing and was imputed by ETS after the assessment. The resulting variable DSEX contains a value for every student and is used for gender comparisons among students.

DRACE7 (Race/Ethnicity for Cross-Sectional)

The variable DRACE7 is an imputed definition of race/ethnicity, derived from up to three sources of information. This variable is used for race/ethnicity subgroup comparisons for the 1994 Cross-sectional assessments in reading, U.S. history, and geography. Two items from the student demographics questionnaire were used in the determination of derived race/ethnicity:

Demographic Item Number 2:

2. If you are Hispanic, what is your Hispanic background?

○ I am not Hispanic.
○ Mexican, Mexican American, or Chicano
○ Puerto Rican
○ Cuban
○ Other Spanish or Hispanic background

Students who responded to item number 2 by filling in the second, third, fourth, or fifth oval were considered Hispanic. For students who filled in the first oval, did not respond to the item, or provided information that was illegible or could not be classified, responses to item number 1 were examined in an effort to determine race/ethnicity. Item number 1 read as follows:

Demographic Item Number 1:

1. Which best describes you?

   ○ White (not Hispanic)

   ○ Black (not Hispanic)

   ○ Hispanic ("Hispanic" means someone who is Mexican, Mexican American, Chicano, Puerto Rican, Cuban, or other Spanish or Hispanic background.)

   ○ Asian ("Asian" means someone who is Chinese, Japanese, Korean, Vietnamese, or other Asian background.)

   ○ Pacific Islander ("Pacific Islander" means someone who is from a Filipino, Hawaiian, or other Pacific Island background.)

   ○ American Indian or Alaskan Native ("American Indian or Alaskan Native" means someone who is from one of the American Indian tribes, or one of the original people of Alaska.)

   ○ Other

Students' race/ethnicity was then assigned to correspond with their selection. For students who filled in the seventh oval (Other), provided illegible information or information that could not be classified, or did not respond at all, race/ethnicity as provided from school records was used.

Derived race/ethnicity could not be determined for students who did not respond to background items 1 or 2 and for whom race/ethnicity was not provided by the school.

DRACE (Self-Identified Race/Ethnicity for Long-Term Trend)

The variable DRACE is an imputed definition of race/ethnicity, derived primarily from student self-identification. This variable is used for race/ethnicity subgroup comparisons in the 1994 trend samples (reading, math, and science) and in the 1994 writing trend samples (see also "Observed race/ethnicity" below). Two items from the student demographics questionnaire were used in the determination of self-identified race/ethnicity:

Demographic Item Number 2:

2. If you are Hispanic, what is your Hispanic background?

   - ○ I am not Hispanic.
   - ○ Mexican, Mexican American, or Chicano
   - ○ Puerto Rican
   - ○ Cuban
   - ○ Other Spanish or Hispanic background

Students who responded to item number 2 by filling in the second, third, fourth, or fifth oval were considered Hispanic. For students who filled in the first oval, did not respond to the item, or provided information that was illegible or could not be classified, responses to item number 1 were examined in an effort to determine race/ethnicity. Item number 1 read as follows:

Demographic Item Number 1:

1. Which best describes you?

   - ○ White (not Hispanic)

   - ○ Black (not Hispanic)

   - ○ Hispanic ("Hispanic" means someone who is Mexican, Mexican American, Chicano, Puerto Rican, Cuban, or from some other Spanish or Hispanic background.)

   - ○ Asian or Pacific Islander ("Asian or Pacific Islander" means someone who is Chinese, Japanese, Korean, Filipino, Vietnamese, or from some other Asian or Pacific Island background.)

   - ○ American Indian or Alaskan Native ("American Indian or Alaskan Native" means someone who is from one of the American Indian tribes, or one of the original people of Alaska.)

   - ○ Other (What?) _________

Students' race/ethnicity was then assigned to correspond with their selection. For students who filled in the sixth oval (Other), provided illegible information or information that could not be classified, or did not respond at all, race/ethnicity as provided from school records was used.

Race/ethnicity could not be determined for students who did not respond to background items 1 or 2 and for whom race/ethnicity was not provided by the school.

RACE (Observed Race/Ethnicity)

The variable RACE is the race/ethnicity of the student being assessed as observed and recorded by the exercise administrator. Observed race/ethnicity was used in NAEP assessments before 1984 (see also "Self-identified race/ethnicity" above). Observed race/ethnicity is used for all race/ethnicity subgroup trend comparisons for which the starting trend point comes from pre-1984 assessment data. Table B-1 shows which of the race/ethnicity variables (observed or self-identified) were used for each subject area report for the 1994 assessment.

Table B-1
Race/Ethnicity Definitions Used in NAEP Reports for the 1994 Assessment

Report	Race/Ethnicity Definition Used
NAEP 1994 Reading Report Card for the Nation and the States: Findings from the National Assessment of Educational Progress and Trial State Assessments	Self-identified (DRACE7)
NAEP 1994 U.S. History Report Card: Findings from the National Assessment of Educational Progress	Self-identified (DRACE7)
NAEP 1994 Geography Report Card: Findings from the National Assessment of Educational Progress	Self-identified (DRACE7)
NAEP 1994 Trends in Academic Progress: Achievement of U.S. Students in Science, 1969 to 1994; Mathematics, 1973 to 1994; Reading, 1971 to 1994; Writing, 1984 to 1994	Self-identified for writing (DRACE); observed for reading, mathematics, and science

TOL8 (Type of Location)
TOL5
TOL3

The type of location variable was new for NAEP in 1994. It is closely based on a similar variable used by NCES in its Common Core of Data Public School Universe. This variable was used for three reasons. First, it seemed desirable to be consistent conceptually with other NCES data products. Second, the necessary data were available for each school to implement the code. Third, the classification is detailed (there are eight levels in all), thus giving maximum information and flexibility in reporting. The levels for type of location are:

1	Large Central City	a central city of a Metropolitan Statistical Area (MSA) or Primary Metropolitan Statistical Area (PMSA)with a population greater than or equal to 400,000, or a population density greater than or equal to 6,000 persons per square mile.

2	Midsize Central City	a central city of an MSA/PMSA, but not a Large Central City
3	Urban Fringe of Large Central City	a place defined as urban that is within an MSA/PMSA that contains a Large Central City
4	Urban Fringe of Midsize City	a place defined as urban within an MSA/PMSA that contains no Large Central City
5	Large Town	a place not within an MSA/PMSA, with a population greater than or equal to 25,000
6	Small Town	a place not within an MSA, with a population less than 25,000, but greater than or equal to 2,500
7	Rural MSA	a place defined as rural (i.e., not within an Urbanized Area) within an MSA/PMSA
8	Rural NonMSA	a place not within an MSA/PMSA with a population of less than 2,500

These categories are designed to be exhaustive and mutually exclusive. Every place in the 50 United States and the District of Columbia is classified as belonging to exactly one of these categories. The definitions of MSAs and PMSAs, and their central cities, is carried out by the Office of Management and Budget (OMB). OMB Bulletin No. 93-17 states that "all agencies that conduct statistical activities to collect and publish data for Metropolitan Areas should use the most recent definitions of Metropolitan Areas established by OMB." The definitions used (as of June 30, 1993) were those current at the time of the 1994 assessment. The definitions of places and their populations are obtained from the published results of the 1990 Population Census, as are the definitions of Urbanized Areas.

Further details about the creation of the eight-category type of location variable are provided in *The NAEP 1994 Sampling and Weighting Report* (Wallace & Rust, 1996).

The variable TOL5 was created by collapsing the information provided in the variable TOL8 to five levels:

1 Large Central City
2 Midsize Central City
3 Urban Fringe of Large City, Urban Fringe of Midsize City, and Large Town
4 Small Town
5 Rural MSA and Rural NonMSA

The variable TOL3 is used extensively in the NAEP reports. TOL3 collapses TOL8 to three levels:

1	Central City	(Large Central City and Midsize Central City) This category includes central cities of all MSAs. Central City is a geographic term and is not synonymous with "inner city."

2	Urban Fringe/Large Town	(Urban Fringe of Large City, Urban Fringe of Midsize City, and Large Town) An Urban Fringe includes all densely settled places and areas within MSAs that are classified as urban by the Bureau of the Census. A Large Town is defined as a place outside MSAs with a population greater than or equal to 25,000.
3	Rural/Small Town	(Small Town, Rural MSA, and Rural NonMSA) Rural includes all places and areas with a population of less than 2,500 that are classified as rural by the Bureau of the Census. A Small Town is defined as a place outside MSAs with a population of less than 25,000 but greater than or equal to 2,500.

PARED (Student's Report of Parents' Education Level)

The variable PARED is derived from responses to two questions, B003501 and B003601, in the student demographic questionnaire. Students were asked to indicate the extent of their mother's education (B003501—How far in high school did your mother go?) by choosing one of the following:

○ She did not finish high school.
○ She graduated from high school.
○ She had some education after high school.
○ She graduated from college.
○ I don't know.

Students were asked to provide the same information about the extent of their father's education (B003601—How far in high school did your father go?) by choosing one of the following:

○ He did not finish high school.
○ He graduated from high school.
○ He had some education after high school.
○ He graduated from college.
○ I don't know.

The information was combined into one parental education reporting category (PARED) as follows: If a student indicated the extent of education for only one parent, that level was included in the data. If a student indicated the extent of education for both parents, the higher of the two levels was included in the data. For students who did not know the level of education for both parents or did not know the level of education for one parent and did not respond for the other, the parental education level was classified as unknown. If the student did not respond for both parents, the student was recorded as having provided no response.

REGION (Region of the Country)

States were grouped into four geographical regions—Northeast, Southeast, Central, and West—as shown in Table B-1. All 50 states and the District of Columbia are listed. The part of Virginia that is included in the Washington, DC, metropolitan statistical area is included in the Northeast region; the remainder of the state is included in the Southeast region.

Table B-1
NAEP Geographic Regions

NORTHEAST	SOUTHEAST	CENTRAL	WEST
Connecticut	Alabama	Illinois	Alaska
Delaware	Arkansas	Indiana	Arizona
District of Columbia	Florida	Iowa	California
Maine	Georgia	Kansas	Colorado
Maryland	Kentucky	Michigan	Hawaii
Massachusetts	Louisiana	Minnesota	Idaho
New Hampshire	Mississippi	Missouri	Montana
New Jersey	North Carolina	Nebraska	Nevada
New York	South Carolina	North Dakota	New Mexico
Pennsylvania	Tennessee	Ohio	Oklahoma
Rhode Island	Virginia	South Dakota	Oregon
Vermont	West Virginia	Wisconsin	Texas
Virginia			Utah
			Washington
			Wyoming

MODAGE (Modal Age)

The modal age (the age of most of the students in the grade sample) for the fourth grade students is age 9. A value of 1 for MODAGE indicates that the student is younger than the modal age; a value of 2 indicates that the student is of the modal age; a value of 3 indicates that the student is older than the modal age.

VARIABLES DERIVED FROM THE STUDENT AND TEACHER QUESTIONNAIRES

Several variables were formed from the systematic combination of response values for one or more items from either the student demographic questionnaire, the student reading background questionnaire, or the teacher questionnaire.

HOMEEN2 (Home Environment—Articles [of 4] in the Home)

The variable HOMEEN2 was created from the responses to student demographic items B000901 (Does your family get a newspaper regularly?), B000903 (Is there an encyclopedia in your home?), B000904 (Are there more than 25 books in your home?), and B000905 (Does your family get any magazines regularly?). The values for this variable were derived as follows:

1	0-2 types	The student responded to at least two items and answered "yes" to two or fewer.
2	3 types	The student answered "yes" to three items.
3	4 types	The student answered "yes" to four items.
8	Omitted	The student answered fewer than two items.

SINGLEP (How Many Parents Live at Home)

SINGLEP was created from items B005601 (Does either your mother or your stepmother live at home with you?) and B005701 (Does either your father or your stepfather live at home with you?). The values for SINGLEP were derived as follows:

1	2 parents at home	The student answered "yes" to both items.
2	1 parent at home	The student answered "yes" to B005601 and "no" to B005701, or "yes" to B005701 and "no" to B005601.
3	Neither at home	The student answered "no" to both items.
8	Omitted	The student did not respond to or filled in more than one oval for one or both items.

TRUMAJ (Teacher Undergraduate Major - Reading)

Items T040701 and T040705 through T040710 in the teacher questionnaire (What were your undergraduate major fields of study?) were used to determine TRUMAJ as follows:

1	English/Reading	The teacher responded "yes" to T040706 or T040707 (English, reading, and/or language arts).
2	Education	The teacher responded "yes" to T040701 (education) and "no" to T040706 and T040707.
3	Other	Any other response.

TRGMAJ (Teacher Graduate Major - Reading)

Items T040801 and T040805 through T040811 in the teacher questionnaire (What were your graduate major fields of study?) were used to determine TRGMAJ as follows:

1	English/Reading	The teacher responded "yes" to T040807 or T040808 (English, reading, and/or language arts).
2	Education	The teacher responded "yes" to T040801 (education) and "no" to T040807 and T040808.
3	Other	The teacher responded "yes" to T040805 (other), T040809 (geography), T040810 (U.S. history), or T040811 (social studies).
4	None	The teacher indicated (T040806) that he or she had no graduate-level study.

HISYRS2 (Number of Years of U.S. History Taken - 8th Graders)

Items H803301, H803302, and H803401 in the U.S. history background section asked students to indicate if they had taken a U.S. history course in 6th grade and in 7th grade, and if they are currently taking a U.S. history course. The variable HISYRS2 is defined as the sum (0-3) of the number of years that the student indicated he/she had (was) taking a U.S. history course.

1	Zero	Did not indicate taking any U.S. history course
2	One	Took (taking) one U.S. history course
3	Two	Took (taking) two U.S. history courses
4	Three	Took U.S. history course in 6th and 7th grade and currently taking U.S. history course

HISYRX2 (Number of Years of U.S. History Taken - 8th Graders - Version 2)

HISYRX2 is defined to be the same as HISYRS2 except that students were assigned a value of "I dont't know" if they did not know if they had (were) taking a U.S. history course for one or more of the variables.

HISYRS3 (Number of Years of U.S. History Taken - 12th Graders)

Items H803303, H803304, H803305, and H803401 in the U.S. history background section asked students to indicate if they had taken a U.S. history course in 9th grade, 10th grade, and in 11th grade, and if they are currently taking a U.S. history course. The variable HISYRS3 is defined as the sum (0-4) of the number of years that the student indicated he/she had (was) taking a U.S. history course.

1	Zero	Did not indicate taking any U.S. history course
2	One	Took (taking) one U.S. history course
3	Two	Took (taking) two U.S. history courses
4	Three	Took (taking) three U.S. history courses
5	Four	Took U.S. history course in 9th, 10th, and 11th grade and currently taking U.S. history course

HISYRX3 (Number of Years of U.S. History Taken - 12th Graders - Version 2)

HISYRX3 is defined to be the same as HISYRS3 except that students were assigned a value of "I dont't know" if they did not know if they had (were) taking a U.S. history course for one or more of the variables.

HISHWK (Amount of U.S. History Homework for Students Who Are Currently Taking History)

HISHWK is defined only for students who indicated that they are currently taking a U.S. history course (H803401 = "yes"). The values of HISHWK are identical to H803501 for these students.

THUMAJ (Teacher Undergraduate Major - U.S. History)

Items T040708, T040709, T040710, T040701, and T040705 in the teacher questionnaire (What were your undergraduate major fields of study?) were used to determine THUMAJ as follows:

1	History/Geog/ Social Studies	The teacher responded "yes" to T040708, T040709, or T040710 (U.S. history, geography, or social studies education).
2	Education	The teacher responded "yes" to T040701 (education) and "no" to T040708, T040709 and T040710.
3	Other	Any other response.

THGMAJ (Teacher Graduate Major - U.S. History)

Items T040806, T040809, T040810, T040811, T040801, and T040805 (What were your graduate major fields of study?) were used to determine TRGMAJ as follows:

1	History/Geog/ Social Studies	The teacher responded "yes" to T040806, T040809, or T040810 (U.S. history, geography, or social studies education).
2	Education	The teacher responded "yes" to T040811 (education) and "no" to T040806, T040809 and T040810.
3	Other	The teacher responded "yes" to T040805 (other)
4	None	The teacher indicated (T040806) that he or she had had no graduate-level study.

THUMAJ2 (Teacher Undergraduate Major in U.S. History)

Items T040708, T040709, T040710, T040701, and T040705 in the teacher questionnaire (What were your undergraduate major fields of study?) were used to determine THUMAJ2 as follows:

1	Yes	The teacher responded "yes" to T040708, T040709, or T040710 (U.S. history, geography, or social studies education).
2	No	Any other response.

THGMAJ2 (Teacher Graduate Major in U.S. History)

Items T040806, T040809, T040810, T040811, T040801, and T040805 in the teacher questionnaire (What were your graduate major fields of study?) were used to determine THGAJ2 as follows:

1	Yes	The teacher responded "yes" to T040806, T040809, or T040810 (U.S. history, geography, or social studies education).
2	No	Any other response.

GEOTAKE (Have Taken at Least One Geography Class)

Items G801601 - G801603 ask students to indicate if they have taken (are taking) a geography class in 6th, 7th, or 8th grade. Similarly, items G801604 - G801607 ask students to indicate if they have taken (are taking) a geography class in 9th, 10th, 11th, or 12th grade. The variable GEOTAKE is defined as:

1	Yes	Student indicated that they have taken (or are currently taking) one or more geography courses.
2	No	Student indicated that they did not take any geography courses
8	Omitted	Student indicated that they did not know or they did not respond

TUNDGS (Teacher Undergraduate Major in Geography/Social Studies)

Items T040708, T040709, T040710, T040701, and T040705 in the teacher questionnaire (What were your undergraduate major fields of study?) were used to determine TUNDGS as follows:

1	Yes	The teacher responded "yes" to T040709 or T040710 (geography or social studies education).
2	No	Any other response.

TUNDGS2 (Teacher Undergraduate Major in Geography/Social Studies/U.S. History)

Items T040708, T040709, T040710, T040701, and T040705 in the teacher questionnaire (What were your undergraduate major fields of study?) were used to determine TUNDGS2 as follows:

1	Yes	The teacher responded "yes" to T040709, T040708, or T040710 (geography, social studies education, or U.S. history).
2	No	Any other response.

TGNDGS (Teacher Graduate Major in Geography/Social Studies)

Items T040806, T040809, T040810, T040805, and T040801 in the teacher questionnaire (What were your graduate major fields of study?) were used to determine TGNDGS as follows:

1	Yes	The teacher responded "yes" to T040809 or T040810 (geography or social studies education).
2	No	Any other response.

TGNDGS2 (Teacher Graduate Major in Geography/Social Studies/U.S. History)

Items T040806, T040809, T040810, T040805, and T040801 in the teacher questionnaire (What were your graduate major fields of study?) were used to determine TGNDGS as follows:

1	Yes	The teacher responded "yes" to T040809, T040810, or T040811 (geography, social studies education, or U.S. history).
2	No	Any other response.

NCOMP (Number of Computer Science Courses Taken)

For age class 17 (long-term trend math and science), NCOMP was created from responses to items B005312 and B005313 concerning the student's coursework in computer science. The values for NCOMP were derived as follows:

1	0	The student answered HAVE NOT to both courses.
2	1	The student answered HAVE to one course.
3	2	The student answered "yes" to both courses.
8	No response	The student did not respond to one or both items.
9	Mult. & out-of-range	The student filled in more than one oval for both items.

NMATH (Highest Level of Mathematics Courses Taken)

For age class 17, long-term trend math and science) NMATH was created from responses to items B005301 through B005307 concerning the student's coursework in mathematics. The values for NMATH were derived as follows:

1	Gen. math or pre-algebra	The student answered HAVE NOT to all items or HAVE to B005301 or B005302 and HAVE NOT to all others.
2	Algebra	The student answered HAVE to B005303 and HAVE NOT to B005304, B005305, B005306, and B005307.
3	Geometry	The student answered HAVE to both B005303 and B005305 and HAVE NOT to B005304, B005306, and B005307.
4	Algebra 2	The student answered HAVE to B005304 or B005306 but HAVE NOT to B005307.
5	Calculus	The student answered HAVE to B005307.
6	Something else	Any other response combination
8	No response	The student did not respond to any item.

NSCI (Highest Level of Science Courses Taken)

For age class 17,(long-term trend math and science) NSCI was created from responses to items B005308 through B005311, which concerned the student's coursework in science. The values for NSCI were derived as follows:

1	No biology	The student answered HAVE NOT to all items or HAVE to B005308 and other than HAVE to all other items.
2	Biology	The student answered HAVE to B005309 and other than HAVE to both B005310 and B005311.
3	Chemistry	The student answered HAVE to both B005309 and B005310 and other than HAVE to B005311.
4	Physics	The student answered HAVE to B005309, B005310, and B005311.
5	Something else	Any other response combination
8	No response	The student answered none of the items.

VARIABLES DERIVED FROM COGNITIVE ITEMS

BKSCOR (Booklet-Level Score)

The booklet level score is a student-level score based on the sum of the number correct for dichotomous items plus the sum of the scores on the polytomous items, where the score for a polytomous item starts from 0 for the unacceptable category. Thus, for a 4-point extended constructed-response item, scores of "no response", "off-task", and "unsatisfactory" are assigned an item score of 0. Scores of "partial", "essential", and "extensive" are assigned item scores of 1, 2, and 3, respectively. The score is computed based on all cognitive items in an individual's assessment booklet.

LOGIT (Logit Percent Correct Within Booklet)

In order to compute the LOGIT score, a percent correct within booklet was first computed. This score was based on the ratio of the booklet score (BKSCOR) over the maximum booklet score. The percent correct score was set to .0001 if no items were answered correctly; if BKSCOR equaled the maximum booklet score, the percent correct score was set to .9999. A logit score, LOGIT, was calculate for each student by the following formula:

A logit score, LOGIT, was calculated within booklet for each student by the following formula:

$$\mathrm{LOGITP} = \ln\left[\frac{PCTCOR}{1 - PCTCOR}\right]$$

LOGIT was then truncated to a value x, such that $-3 \leq x \leq 3$. After computing LOGIT for each student, the mean and standard deviation was calculated for each booklet as the first step in standardizing the logit scores. The standardized logit score, ZLOGIT, was then calculated for each student by the following formula:

$$\text{ZLOGIT} = \left[\frac{\mathit{LOGIT} - \mathit{mean\ logit}}{\mathit{standard\ deviation}}\right]$$

NORMIT (Normit Gaussian Score)
SCHNORM (School-Level Mean Gaussian Score)

The normit score is a student-level Gaussian score based on the inverse normal transformation of the mid-percentile rank of a student's number-correct booklet score within that booklet. The normit scores were used to decide collapsing of variables, finalize conditioning coding, and check the results of scaling.

The number correct is based on the number of dichotomous items answered correctly plus the score obtained on extended constructed-response items. The mid-percentile rank is based on the formula:

$$\frac{CF(i) + CF(i-1)}{2N}$$

where CF(I) is the cumulative frequency at I items correct and N is the total sample size. If I = 0 then

$$\frac{CF(0) + \frac{CF(1)}{2}}{2N}$$

A school-level normit, SCHNORM, was also created; this was the mean normit across all reading booklets administered in a school. These school-level mean normit scores were used in conditioning procedures to take into account differences in school proficiency. For each school, the weighted mean of the logits for the students in that school was calculated. Each student was then assigned that mean as his or her school-level mean logit score value.

VARIABLES RELATED TO PROFICIENCY SCALING

Proficiency Score Variables

Item response theory (IRT) was used to estimate average proficiency for the nation and for various subpopulations, based on students' performance on the set of cognitive items they received. IRT provides a common scale on which performance can be reported for the nation and subpopulations, even when all students do not answer the same set of questions. This common scale makes it possible to report on relationships between students' characteristics (based on their responses to the background questions) and their overall performance in the assessment.

A scale ranging from 1 to 500 was created to report performance for each content area. A composite scale was created based on a weighted average of subscales, where the weight for each content area was proportional to the relative importance assigned to the content area as specified in the reading, history, and geography objectives.

Scale proficiency estimates were obtained for all students. The NAEP methods use random draws (plausible values) from estimated proficiency distributions to compute population statistics. Plausible values are not optimal estimates of individual proficiency; instead, they serve as intermediate values to be used in estimating population characteristics. Chapter 11 provides further details on the computation and use of plausible values. Chapters 12-18 provide additional information as appropriate to each sample/subject area.

The proficiency score (plausible value) variables are provided on the student data files for each of the scales and are named as shown in Table B-2.

Table B-2
Scaling Variables for the 1994 National Assessment Samples

Sample	Scale	Data Variables
Reading Cross-Sectional (Main)	Reading for Literary Experience	RRPS11 to RRPS15
	Reading to Gain Information	RRPS21 to RRPS25
	Reading to Perform a Task	RRPS31 to RRPS35
	Composite	RRPCM1 to RRPCM5
U.S. History Cross-Sectional (Main)	Democracy	HRPS11 to HRPS15
	Cultures	HRPS21 to HRPS25
	Technology	HRPS31 to HRPS35
	World Role	HRPS41 to HRPS45
	Composite	HRPCM1 to HRPCM5
Geography Cross-Sectional (Main)	Space and Place	GRPS11 to GRPS51
	Environment and Society	GRPS21 to GRPS25
	Spatial Dynamics and Connections	GRPS31 to GRPS35
	Composite	GRPCM1 to GRMCP5
Reading Long-Term Trend	Univariate	REDVAL1 to REDVAL5
Writing Long-Term Trend	Univariate	WRPSCT1 to WRPSCT5
Mathematics Long-Term Trend	Univariate	MRPSCT1 to MRPSCT5
Science Long-Term Trend	Univariate	SRPSCT1 to SRPSCT5

SMEANx, SMNx1	**(School Mean Score Using First Plausible Value)**
SRANKx, SRNKx1	**(School Rank Using First Plausible Value)**
SRNK3x, SRK3x1	**(Top, Middle, Bottom Third Using First Plausible Value**
where x = R, H, or G	**for Reading, U.S. History, or Geography)**

A mean reading composite score (SMEANx on the student files, SMNx1 on the school files) was calculated using the first composite plausible value for each school within each grade/subject area. The mean composite score was based on the values from the scaling variable xRPCM1 and was calculated using the students' sampling weights. The schools were then ordered from highest to lowest mean score (SRANKx on the student files, SRNKx1 on the school files) within a sample using school-level weights—the school with the highest mean score was given a ranking of 1 and the school with the lowest mean score was given a ranking equal to the number of schools in the sample.

These variables were then used in partitioning the schools within the national grade sample into three groups (top third, middle third, and bottom third) based on their ranking (SRNK3x on the student files, SRK3x1 on the school files).

SMEANRP, SMNR1P	**(School Mean Score Using First Plausible Value, Public Schools Only)**
SRANKRP, SRNKR1P	**(School Rank Using First Plausible Value, Public Schools Only)**
SRNK3RP, SRK3R1P	**(Top, Middle, Bottom Third, Using First Plausible Value, Public Schools Only)**

These variables were computed in the same manner as SMEANR, SMNR1, SRANKR, SRNKR1, SRNK3R, and SRK3R1 for the subset of students who attended public schools. Note that this set of variables is available only for grade 4 reading.

SMNxn	**(School Mean Score Using Plausible Values 2 Through 5)**
SRNKxn	**(School Rank Using Plausible Values 2 Through 5)**
SRK3xn	**(Top, Middle, Bottom Third Using Plausible Values 2 Through 5)**
SMNRPn	**(School Mean Score Using Plausible Values 2 Through 5, Public Schools Only)**
SRNKRPn	**(School Rank Using Plausible Values 2 Through 5, Public Schools Only)**
SRK3RPn	**(Top, Middle, Bottom Third, Using Plausible Values 2 Through 5, Public Schools Only)**

School ranking results presented in the 1994 NAEP reports are based on the first plausible value. However, since there are four additional estimates of proficiency (plausible values) for each student, school ranking data were also created for those estimates. These school rank values were created using the same procedures described above, substituting proficiency variables xRPCM2 through xRPCM5 to compute the results. In the variable names, n denotes the plausible value 2, 3, 4, or 5. Note that these variables are included only on the school file. Note that the variables SMNRPn, SRNKRPn, and SRK3RPn are available only for grade 4 reading.

QUALITY EDUCATION DATA VARIABLES (QED)

The data files contain several variables obtained from information supplied by Quality Education Data, Inc. (QED). QED maintains and updates annually lists of schools showing grade span, total enrollment, instructional dollars per pupil, and other information for each school. These data variables are retained on both the school and student files and are identified in the data layouts by "QED" in the SHORT LABEL field.

Most of the QED variables are defined sufficiently in the data codebooks. Explanations of others are provided below.

ORSHPT and SORSHPT are the Orshansky Percentile, an indicator of relative wealth that specifies the percentage of school-age children in a district who fall below the poverty line.

IDP and SIDP represent, at the school district level, dollars per student spent for textbooks and supplemental materials.

ADULTED and SADLTED indicate whether or not adult education courses are offered at the school site.

URBAN and SURBAN define the school's urbanicity: urban (central city); suburban (area surrounding central city, but still located within the counties constituting the metropolitan statistical area); or rural (area outside any metropolitan statistical area).

Appendix C

CONDITIONING VARIABLES AND CONTRAST CODINGS

Appendix C

CONDITIONING VARIABLES AND CONTRAST CODINGS

This appendix contains information about the conditioning variables used in scaling/plausible value estimation for the 1994 NAEP assessment. The initial step in construction of conditioning variables involves forming primary student-based vectors of response data from answers to student, teacher, and school questionnaires, demographic and background data such as supplied by Westat, Inc., and other student information known prior to scaling. The initial conditioning vectors concatenate this student background information into a series of identifying "contrasts" comprising:

1. Categorical variables derived by expanding the response options of a questionnaire variable into a binary series of one-degree-of-freedom "dummy" variables or contrasts, (these form the majority of each student conditioning vector);

2. Questionnaire or demographic variables that possess ordinal response options, such as number of hours spent watching television, which are included as linear and/or quadratic multi-degree-of-freedom contrasts;

3. Continuous variables, such as student logit scores based on percent correct values, included as contrasts in their original form or a transformation of their original form, and;

4. Interactions of two or more categorical variables forming a set of orthogonal one-degree-of-freedom dummy variables or contrasts.

This appendix gives the specifications used for constructing the conditioning variables. Table C-1 defines the information provided for each main sample variable. Conditioning variable data used in common across the main reading, U.S. history, and geography assessments are given in Table C-2; Table C-3 shows additional conditioning variable data used in common for both U.S. history and geography. Conditioning variable data specific to each subject area is shown for main reading in Table C-4, for U.S. history in Table C-5, and for geography in Table C-6. Similar information for long-term trend samples is given in Tables C-7, C-8, C-9, and C-10 respectively for reading, mathematics, science, and writing.

As described in Chapter 9, the linear conditioning model employed for the estimation of plausible values did not directly use the conditioning variable specifications listed in this appendix. To eliminate inherent instabilities in estimation encountered when using a large number of correlated variables, a principal component transformation of the correlation matrix obtained from the conditioning variable contrasts derived according to these primary specifications was performed. The principal components scores based on this transformation were used as the predictor variables in estimating the linear conditioning model. The proportions of variance of the conditioning contrast accounted for by the principal components are given for each age/grade level in Tables C-11, C-12, and C-13 for reading; Tables C-14, C-15, and C-16 for U.S. history, and Tables C-17, C-18, and C-19 for geography.

Table C-1

Description of Specifications Provided for Each Conditioning Variable

Title	Description
CONDITIONING ID	An unique eight-character ID assigned to identify each conditioning variable corresponding to a particular background or subject area question within the entire pool of conditioning variables. The first four characters identify the origin of the variable: BACK (background questionnaire), READ (student reading questionnaire), SCHL (school questionnaire), TCHR (background part of teacher questionnaire), and TSUB (subject classroom part of teacher questionnaire). The second four digits represent the sequential position within each origin group.
DESCRIPTION	A short description of the conditioning variable.
GRADES/ASSESSMENTS	Three characters identifying assessment ("S" for state, "N" for national) and grade (04, 08, and 12) in which the conditioning variable was used.
CONDITIONING VAR LABEL	A descriptive eight-character label identifying the conditioning variable.
NAEP ID	The seven-character NAEP database identification for the conditioning variable.
TYPE OF CONTRAST	The type of conditioning variable. "CLASS" identifies a categorical conditioning variable and "SCALE" identifies continuous or quasi-continuous conditioning variables. "INTERACTION" identifies a set of orthogonal contrasts formed from two or more "CLASS" variables. "OTHER" conditioning variables do not fall into any of the above types.
TOTAL NUMBER OF SPECIFIED CONTRASTS	Each conditioning variable forms a set of one or more contrasts. For each valid response value of conditioning variable a contrast must be defined. One or more response values may be collapsed together to form one contrast. The number of response value "sets" of a conditioning variable forming a unique contrast is the value given in this field.
NUMBER OF INDEPENDENT CONTRASTS	The number of degree of freedom in a set of contrasts formed from a conditioning variable. For a categorical conditioning variable this number would be the number of response options minus one if each response option formed its own unique contrast.

Table C-2
Conditioning Variables Common to the Main Reading, U.S. History, and Geography Assessments

```
CONDITIONING VARIABLE ID:  BACK0001
DESCRIPTION:               GRAND MEAN
GRADES/ASSESSMENTS:        N04, S04, N08, N12
CONDITIONING VAR LABEL:    OVERALL
NAEP ID:                   BKSER                  TOTAL NUMBER OF SPECIFIED CONTRASTS:     1
TYPE OF CONTRAST:          OTHER                  NUMBER OF INDEPENDENT CONTRASTS:         1

001 OVERALL  (@          ) 1                      GRAND MEAN

CONDITIONING VARIABLE ID:  BACK0002
DESCRIPTION:               DERIVED SEX
GRADES/ASSESSMENTS:        N04, S04, N08, N12
CONDITIONING VAR LABEL:    GENDER
NAEP ID:                   DSEX                   TOTAL NUMBER OF SPECIFIED CONTRASTS:     2
TYPE OF CONTRAST:          CLASS                  NUMBER OF INDEPENDENT CONTRASTS:         1

001 MALE     (1          ) 0                      MALE
002 FEMALE   (2          ) 1                      FEMALE

CONDITIONING VARIABLE ID:  BACK0003
DESCRIPTION:               DERIVED RACE/ETHNICITY
GRADES/ASSESSMENTS:        N04, S04, N08, N12
CONDITIONING VAR LABEL:    RACE/ETH
NAEP ID:                   DRACE7                 TOTAL NUMBER OF SPECIFIED CONTRASTS:     5
TYPE OF CONTRAST:          CLASS                  NUMBER OF INDEPENDENT CONTRASTS:         4

001 WHI/AI/O (1,6,7      ) 0000                   RACE/ETHNICITY:  WHITE, AMERICAN INDIAN/ALASKAN
                                                  NATIVE, OTHER, MISSING, UNCLASSIFIED
002 BLACK    (2          ) 1000                   RACE/ETHNICITY:  BLACK
003 HISPANIC (3          ) 0100                   RACE/ETHNICITY:  HISPANIC
004 ASIAN    (4          ) 0010                   RACE/ETHNICITY:  ASIAN
005 PAC ISLD (5          ) 0001                   RACE/ETHNICITY:  PACIFIC ISLANDER

CONDITIONING VARIABLE ID:  BACK0004
DESCRIPTION:               IF HISPANIC, WHAT IS YOUR HISPANIC BACKGROUND?
GRADES/ASSESSMENTS:        N04, S04, N08, N12
CONDITIONING VAR LABEL:    HISPANIC
NAEP ID:                   B003101                TOTAL NUMBER OF SPECIFIED CONTRASTS:     5
TYPE OF CONTRAST:          CLASS                  NUMBER OF INDEPENDENT CONTRASTS:         4

001 NOT HISP (1          ) 0000                   HISPANIC:  NOT HISPANIC
002 MEXICAN  (2          ) 1000                   HISPANIC:  MEXICAN, MEXICAN AMERICAN, CHICANO
003 PUER RIC (3          ) 0100                   HISPANIC:  PUERTO RICAN
004 CUBN,OTH (4,5        ) 0010                   HISPANIC:  CUBAN, OTHER
005 HISP-?   (M          ) 0001                   HISPANIC:  MISSING

CONDITIONING VARIABLE ID:  BACK0005
DESCRIPTION:               MSA/NON-MSA
GRADES/ASSESSMENTS:        N04, N08, N12
CONDITIONING VAR LABEL:    MSANAT
NAEP ID:                   TOL8                   TOTAL NUMBER OF SPECIFIED CONTRASTS:     2
TYPE OF CONTRAST:          CLASS                  NUMBER OF INDEPENDENT CONTRASTS:         1

001 MSA      (1,2,3,4,7,M) 0                      MSA
002 NON MSA  (5,6,8      ) 1                      NON-MSA

CONDITIONING VARIABLE ID:  BACK0006
DESCRIPTION:               TYPE OF LOCALE (5 CATEGORIES)
GRADES/ASSESSMENTS:        N04, S04, N08, N12
CONDITIONING VAR LABEL:    TOL5
NAEP ID:                   TOL5                   TOTAL NUMBER OF SPECIFIED CONTRASTS:     5
TYPE OF CONTRAST:          CLASS                  NUMBER OF INDEPENDENT CONTRASTS:         4
```

BEST COPY AVAILABLE

Table C-2 (continued)

Conditioning Variables Common to the Main Reading, U.S. History, and Geography Assessments

```
001 BIG CTY5 (1          ) 0000                    TOL5:  LARGE CITY
002 MID CTY5 (2,M        ) 1000                    TOL5:  MID-SIZE CITY
003 FR/BTWN5 (3          ) 0100                    TOL5:  URBAN FRINGE OF LARGE CITY, URBAN FRINGE OF
                                                          MID-SIZE CITY
004 SML TWN5 (4          ) 0010                    TOL5:  SMALL TOWN
005 RURAL5   (5          ) 0001                    TOL5: -RURAL (MSA AND NON-MSA)

CONDITIONING VARIABLE ID:  BACK0007
DESCRIPTION:               DESCRIPTION OF COMMUNITY
GRADES/ASSESSMENTS:        N04, S04, N08, N12
CONDITIONING VAR LABEL:    DOC
NAEP ID:                   DOC                     TOTAL NUMBER OF SPECIFIED CONTRASTS:     4
TYPE OF CONTRAST:          CLASS                   NUMBER OF INDEPENDENT CONTRASTS:         3

001 BIG CITY (1          ) 000                     DOC:  BIG CITY
002 URBAN FR (2          ) 100                     DOC:  URBAN FRINGE
003 MED CITY (3,M        ) 010                     DOC:  MEDIUM CITY
004 SM PLACE (4          ) 001                     DOC:  SMALL PLACE

CONDITIONING VARIABLE ID:  BACK0008
DESCRIPTION:               PARENTS' HIGHEST LEVEL OF EDUCATION
GRADES/ASSESSMENTS:        N04, S04, N08, N12
CONDITIONING VAR LABEL:    PARED
NAEP ID:                   PARED                   TOTAL NUMBER OF SPECIFIED CONTRASTS:     5
TYPE OF CONTRAST:          CLASS                   NUMBER OF INDEPENDENT CONTRASTS:         4

001 < HS     (1          ) 0000                    PARED:  LESS THAN HIGH SCHOOL
002 HS GRAD  (2          ) 1000                    PARED:  HIGH SCHOOL GRADUATE
003 POST HS  (3          ) 0100                    PARED:  POST HIGH SCHOOL
004 COL GRAD (4          ) 0010                    PARED:  COLLEGE GRADUATE
005 PARED-?  (5,M        ) 0001                    PARED:  MISSING, I DON'T KNOW

CONDITIONING VARIABLE ID:  BACK0009
DESCRIPTION:               REGION OF THE COUNTRY
GRADES/ASSESSMENTS:        N04, N08, N12
CONDITIONING VAR LABEL:    REGION
NAEP ID:                   REGION                  TOTAL NUMBER OF SPECIFIED CONTRASTS:     4
TYPE OF CONTRAST:          CLASS                   NUMBER OF INDEPENDENT CONTRASTS:         3

001 N EAST   (1          ) 000                     REGION:  NORTHEAST
002 S EAST   (2          ) 100                     REGION:  SOUTHEAST
003 CENTRAL  (3          ) 010                     REGION:  CENTRAL
004 WEST     (4,5        ) 001                     REGION:  WEST, TERRITORIES (NONE)

CONDITIONING VARIABLE ID:  BACK0010
DESCRIPTION:               SCHOOL TYPE (PQ)
GRADES/ASSESSMENTS:        N04, S04, N08, N12
CONDITIONING VAR LABEL:    SCHTYPE
NAEP ID:                   SCHTYPE                 TOTAL NUMBER OF SPECIFIED CONTRASTS:     3
TYPE OF CONTRAST:          CLASS                   NUMBER OF INDEPENDENT CONTRASTS:         2

001 PUBLIC   (1          ) 00                      SCHOOL TYPE:  PUBLIC
002 PRIVATE  (2,4,5,M    ) 10                      SCHOOL TYPE:  PRIVATE, BUREAU OF INDIAN AFFAIRS,
                                                                 DEPARTMENT OF DEFENSE, MISSING
003 CATHOLIC (3          ) 01                      SCHOOL TYPE:  CATHOLIC

CONDITIONING VARIABLE ID:  BACK0011
DESCRIPTION:               INDIVIDUALIZED EDUCATION PLAN
GRADES/ASSESSMENTS:        N04, S04, N08, N12
CONDITIONING VAR LABEL:    IEP
NAEP ID:                   IEP                     TOTAL NUMBER OF SPECIFIED CONTRASTS:     2
TYPE OF CONTRAST:          CLASS                   NUMBER OF INDEPENDENT CONTRASTS:         1
```

BEST COPY AVAILABLE

Table C-2 (continued)

Conditioning Variables Common to the Main Reading, U.S. History, and Geography Assessments

```
001 IEP-YES  (1           ) 0                     IEP:  YES
002 IEP-NO   (2           ) 1                     IEP:  NO

CONDITIONING VARIABLE ID:  BACK0012
DESCRIPTION:               LIMITED ENGLISH PROFICIENCY
GRADES/ASSESSMENTS:        N04, S04, N08, N12
CONDITIONING VAR LABEL:    LEP
NAEP ID:                   LEP                    TOTAL NUMBER OF SPECIFIED CONTRASTS:     2
TYPE OF CONTRAST:          CLASS                  NUMBER OF INDEPENDENT CONTRASTS:         1

001 LEP-YES  (1           ) 0                     LEP:  YES
002 LEP-NO   (2           ) 1                     LEP:  NO

CONDITIONING VARIABLE ID:  BACK0013
DESCRIPTION:               CHAPTER 1 (BOOK COVER)
GRADES/ASSESSMENTS:        N04, S04, N08, N12
CONDITIONING VAR LABEL:    CHAPTER1
NAEP ID:                   CHAP1                  TOTAL NUMBER OF SPECIFIED CONTRASTS:     2
TYPE OF CONTRAST:          CLASS                  NUMBER OF INDEPENDENT CONTRASTS:         1

001 CHAP1-Y  (1           ) 0                     CHAPTER 1:  YES
002 CHAP1-N  (2           ) 1                     CHAPTER 1:  NO

CONDITIONING VARIABLE ID:  BACK0014
DESCRIPTION:               PERCENT WHITE STUDENTS IN SCHOOL (FROM QED)
GRADES/ASSESSMENTS:        N04, S04, N08, N12
CONDITIONING VAR LABEL<    PCTWHITE
NAEP ID:                   PCTWHTQ                TOTAL NUMBER OF SPECIFIED CONTRASTS:     3
TYPE OF CONTRAST:          CLASS                  NUMBER OF INDEPENDENT CONTRASTS:         2

001 PREDOM/? (80-110,M    ) 00                    PREDOMINANTLY WHITE, MISSING
002 INTEGRAT (50-79       ) 10                    INTEGRATED
003 MINORITY (0-49        ) 01                    WHITE MINORITY

CONDITIONING VARIABLE ID:  BACK0015
DESCRIPTION:               DO YOU RECEIVE A FREE OR REDUCED-PRICE LUNCH?
GRADES/ASSESSMENTS:        N04, S04, N08, N12
CONDITIONING VAR LABEL:    LUNCH
NAEP ID:                   B008101                TOTAL NUMBER OF SPECIFIED CONTRASTS:     3
TYPE OF CONTRAST:          CLASS                  NUMBER OF INDEPENDENT CONTRASTS:         2

001 FR LUNCH (1           ) 00                    LUNCH PROGRAM:  FREE/REDUCED
002 NO LUNCH (2           ) 10                    LUNCH PROGRAM:  NOT FREE/REDUCED
003 LUNCH-?  (3,M         ) 01                    LUNCH PROGRAM:  I DON'T KNOW, MISSING

CONDITIONING VARIABLE ID:  BACK0016
DESCRIPTION:               MODAL AGE, MODAL GRADE (DERIVED)
GRADES/ASSESSMENTS:        N04, N08, N12
CONDITIONING VAR LABEL:    AGE/GRAD
NAEP ID:                   MODGRAG                TOTAL NUMBER OF SPECIFIED CONTRASTS:     5
TYPE OF CONTRAST:          CLASS                  NUMBER OF INDEPENDENT CONTRASTS:         4

001 <MA/=MG  (1           ) 0000                  LESS THAN MODAL AGE, MODAL GRADE
002 =MA/<MG  (2           ) 1000                  MODAL AGE, LESS THAN MODAL GRADE
003 =MA/=MG  (3           ) 0100                  MODAL AGE, MODAL GRADE, MISSING
004 =MA/>MG  (4           ) 0010                  MODAL AGE, GREATER THAN MODAL GRADE
005 >MA/=MG  (5           ) 0001                  GREATER THAN MODAL AGE, MODAL GRADE

CONDITIONING VARIABLE ID:  BACK0017
DESCRIPTION:               HOW OFTEN DO THE PEOPLE IN YOUR HOME SPEAK A LANGUAGE OTHER THAN ENGLISH?
GRADES/ASSESSMENTS:        N04, S04, N08, N12
CONDITIONING VAR LABEL:    HOMELANG
NAEP ID:                   B003201                TOTAL NUMBER OF SPECIFIED CONTRASTS:     4
TYPE OF CONTRAST:          CLASS                  NUMBER OF INDEPENDENT CONTRASTS:         3
```

BEST COPY AVAILABLE

Table C-2 (continued)

Conditioning Variables Common to the Main Reading, U.S. History, and Geography Assessments

```
001 HL-NEVER (1           ) 000                        HOMELANG:  NEVER
002 HL-SOME  (2           ) 100                        HOMELANG:  SOMETIMES
003 HL-ALWAY (3           ) 010                        HOMELANG:  ALWAYS
004 HL-?     (M           ) 001                        HOMELANG:  MISSING

CONDITIONING VARIABLE ID:  BACK0018
DESCRIPTION:               HOW MUCH TELEVISION DO YOU USUALLY WATCH EACH DAY? (LINEAR)
GRADES/ASSESSMENTS:        N04, S04, N08, N12
CONDITIONING VAR LABEL:    TVWATCHL
NAEP ID:                   B001801                     TOTAL NUMBER OF SPECIFIED CONTRASTS:     7
TYPE OF CONTRAST:          LINEAR                      NUMBER OF INDEPENDENT CONTRASTS:         1

001 TVLIN-0  (1           ) 0                          TV WATCHING (LINEAR) (0 TO 6+ HOURS PER DAY)
002 TVLIN-1  (2           ) 1                          TV WATCHING (LINEAR)
003 TVLIN-2  (3           ) 2                          TV WATCHING (LINEAR)
004 TVLIN-3  (4,M         ) 3                          TV WATCHING (LINEAR)
005 TVLIN-4  (5           ) 4                          TV WATCHING (LINEAR)
006 TVLIN-5  (6           ) 5                          TV WATCHING (LINEAR)
007 TVLIN-6  (7           ) 6                          TV WATCHING (LINEAR)

CONDITIONING VARIABLE ID:  BACK0019
DESCRIPTION:               HOW MUCH TELEVISION DO YOU USUALLY WATCH EACH DAY? (QUADRATIC)
GRADES/ASSESSMENTS:        N04, S04, N08, N12
CONDITIONING VAR LABEL:    TVWATCHQ
NAEP ID:                   B001801                     TOTAL NUMBER OF SPECIFIED CONTRASTS:     1
TYPE OF CONTRAST:          QUADRATIC                   NUMBER OF INDEPENDENT CONTRASTS:         1

001 TV-QUAD  (1-7,M=4     )  1.0 + -2.0*X +  1.0*X**2  TV WATCHING (QUADRATIC)

CONDITIONING VARIABLE ID:  BACK0020
DESCRIPTION:               HOMEWORK ASSIGNED?:  BASED ON TIME SPENT ON HOMEWORK EACH DAY.
GRADES/ASSESSMENTS:        N04, S04, N08, N12
CONDITIONING VAR LABEL:    HWASSIGN
NAEP ID:                   B006601                     TOTAL NUMBER OF SPECIFIED CONTRASTS:     3
TYPE OF CONTRAST:          CLASS                       NUMBER OF INDEPENDENT CONTRASTS:         2

001 HW-MISS  (M           ) 00                         HOMEWORK ASSIGNED?:  MISSING
002 HW-NO    (1           ) 10                         HOMEWORK ASSIGNED?:  NO
003 HW-YES   (2-5         ) 01                         HOMEWORK ASSIGNED?:  YES

CONDITIONING VARIABLE ID:  BACK0021
DESCRIPTION:               HOW MUCH TIME DO YOU USUALLY SPEND ON HOMEWORK EACH DAY? (LINEAR)
GRADES/ASSESSMENTS:        N04, S04, N08, N12
CONDITIONING VAR LABEL:    HOMEWRKL
NAEP ID:                   B006601                     TOTAL NUMBER OF SPECIFIED CONTRASTS:     4
TYPE OF CONTRAST:          LINEAR                      NUMBER OF INDEPENDENT CONTRASTS:         1

001 HWLIN-0  (1,2,M       ) 0                          HOMEWORK (LINEAR):  DON'T HAVE ANY, DON'T DO ANY,
                                                                           MISSING
002 HWLIN-1  (3           ) 1                          HOMEWORK (LINEAR):  1/2 HOUR OR LESS
003 HWLIN-2  (4           ) 2                          HOMEWORK (LINEAR):  1 HOUR
004 HWLIN-3  (5           ) 3                          HOMEWORK (LINEAR):  MORE THAN 1 HOUR

CONDITIONING VARIABLE ID:  BACK0022
DESCRIPTION:               HOW MUCH TIME DO YOU USUALLY SPEND ON HOMEWORK EACH DAY (QUADRATIC)
GRADES/ASSESSMENTS:        N04, S04, N08, N12
CONDITIONING VAR LABEL:    HOMEWRKQ
NAEP ID:                   B006601                     TOTAL NUMBER OF SPECIFIED CONTRASTS:     4
TYPE OF CONTRAST:          SCALE                       NUMBER OF INDEPENDENT CONTRASTS:         1
```

Table C-2 (continued)
Conditioning Variables Common to the Main Reading, U.S. History, and Geography Assessments

```
001 HWQUAD-0 (1,2,M      ) 0                           HOMEWORK (QUADRATIC):  DON'T HAVE ANY, DON'T DO
                                                                             ANY, MISSING
002 HWQUAD-1 (3          ) 1                           HOMEWORK (QUADRATIC):  1/2 HOUR OR LESS
003 HWQUAD-2 (4          ) 4                           HOMEWORK (QUADRATIC):  1 HOUR
004 HWQUAD-3 (5          ) 9                           HOMEWORK (QUADRATIC):  MORE THAN 1 HOUR

CONDITIONING VARIABLE ID:  BACK0023
DESCRIPTION:               NUMBER OF ITEMS IN THE HOME (NEWSPAPER, > 25 BOOKS, ENCYCLOPEDIA, MAGAZINES)
                           (DERIVED)
GRADES/ASSESSMENTS:        N04, S04, N08, N12
CONDITIONING VAR LABEL:    HOMEITMS
NAEP ID:                   HOMEEN2                     TOTAL NUMBER OF SPECIFIED CONTRASTS:     3
TYPE OF CONTRAST:          CLASS                       NUMBER OF INDEPENDENT CONTRASTS:         2

001 HITEM<=2 (1,M        ) 00                          ITEMS IN HOME:  ZERO TO TWO ITEMS, MISSING
002 HITEM=3  (2          ) 10                          ITEMS IN HOME:  THREE ITEMS
003 HITEM=4  (3          ) 01                          ITEMS IN HOME:  FOUR ITEMS

CONDITIONING VARIABLE ID:  BACK0025
DESCRIPTION:               DOES MOTHER OR STEPMOTHER LIVE AT HOME WITH YOU?
GRADES/ASSESSMENTS:        N04, S04, N08, N12
CONDITIONING VAR LABEL:    MOM@HOME
NAEP ID:                   B005601                     TOTAL NUMBER OF SPECIFIED CONTRASTS:     3
TYPE OF CONTRAST:          CLASS                       NUMBER OF INDEPENDENT CONTRASTS:         2

001 MOMHOM-Y (1          ) 00                          MOTHER AT HOME:  YES
002 MOMHOM-N (2          ) 10                          MOTHER AT HOME:  NO
003 MOMHOM-? (M          ) 01                          MOTHER AT HOME:  MISSING

CONDITIONING VARIABLE ID:  BACK0026
DESCRIPTION:               DOES FATHER OR STEPFATHER LIVE AT HOME WITH YOU?
GRADES/ASSESSMENTS:        N04, S04, N08, N12
CONDITIONING VAR LABEL:    DAD@HOME
NAEP ID:                   B005701                     TOTAL NUMBER OF SPECIFIED CONTRASTS:     3
TYPE OF CONTRAST:          CLASS                       NUMBER OF INDEPENDENT CONTRASTS:         2

001 DADHOM-Y (1          ) 00                          FATHER AT HOME:  YES
002 DADHOM-N (2          ) 10                          FATHER AT HOME:  NO
003 DADHOM-? (M          ) 01                          FATHER AT HOME:  MISSING

CONDITIONING VARIABLE ID:  BACK0027
DESCRIPTION:               HOW MANY DAYS OF SCHOOL MISSED LAST MONTH?
GRADES/ASSESSMENTS:        N04, S04, N08, N12
CONDITIONING VAR LABEL:    SCH MISS
NAEP ID:                   S004001                     TOTAL NUMBER OF SPECIFIED CONTRASTS:     2
TYPE OF CONTRAST:          CLASS                       NUMBER OF INDEPENDENT CONTRASTS:         1

001 MISS->2  (3,4,5,M    ) 0                           DAYS OF SCHOOL MISSED:  3-4, 5-10, 10 OR MORE
                                                                              DAYS, MISSING
002 MISS-2<  (1,2        ) 1                           DAYS OF SCHOOL MISSED:  0-1, 2 DAYS

CONDITIONING VARIABLE ID:  BACK0028
DESCRIPTION:               HOW LONG LIVED IN THE UNITED STATES?
GRADES/ASSESSMENTS:        N04, S04, N08, N12
CONDITIONING VAR LABEL:    YRSINUSA
NAEP ID:                   B008001                     TOTAL NUMBER OF SPECIFIED CONTRASTS:     4
TYPE OF CONTRAST:          CLASS                       NUMBER OF INDEPENDENT CONTRASTS:         3

001 USA >5   (1          ) 000                         LIVED IN US MORE THAN 5 YEARS
002 USA 3-5  (2          ) 100                         LIVED IN US 3-5 YEARS
003 USA <3   (3          ) 010                         LIVED IN US LESS THAN 3 YEARS
004 USA-?    (M          ) 001                         LIVED IN US MISSING
```

Table C-2 (continued)

Conditioning Variables Common to the Main Reading, U.S. History, and Geography Assessments

```
CONDITIONING VARIABLE ID:  BACK0029
DESCRIPTION:               HOW MANY GRADES IN THIS STATE? (4TH GRADE)
GRADES/ASSESSMENTS:        N04, S04
CONDITIONING VAR LABEL:    STGRADE4
NAEP ID:                   B007601                    TOTAL NUMBER OF SPECIFIED CONTRASTS:     3
TYPE OF CONTRAST:          CLASS                      NUMBER OF INDEPENDENT CONTRASTS:         2

001 STGRD<1  (1,M        ) 00                         GRADES IN STATE:  LESS THAN 1 GRADE, MISSING
002 STGRD1-2 (2          ) 10                         GRADES IN STATE:  1-2 GRADES
003 STGRD3>  (3          ) 01                         GRADES IN STATE:  3 OR MORE GRADES

CONDITIONING VARIABLE ID:  BACK0030
DESCRIPTION:               HOW MANY GRADES IN THIS STATE? (8TH GRADE)
GRADES/ASSESSMENTS:        N08
CONDITIONING VAR LABEL:    STGRADE8
NAEP ID:                   M810801                    TOTAL NUMBER OF SPECIFIED CONTRASTS:     4
TYPE OF CONTRAST:          CLASS                      NUMBER OF INDEPENDENT CONTRASTS:         3

001 STGRD<1  (1,M        ) 000                        GRADES IN STATE:  LESS THAN 1 GRADE, MISSING
002 STGRD1-2 (2          ) 100                        GRADES IN STATE:  1-2 GRADES
003 STGRD3-5 (3          ) 010                        GRADES IN STATE:  3-5 GRADES
004 STGRD6>  (4          ) 001                        GRADES IN STATE:  MORE THAN 5 GRADES

CONDITIONING VARIABLE ID:  BACK0031
DESCRIPTION:               HOW MANY GRADES IN THIS STATE (12TH GRADE)
GRADES/ASSESSMENTS:        N12
CONDITIONING VAR LABEL:    STGRAD12
NAEP ID:                   B008301                    TOTAL NUMBER OF SPECIFIED CONTRASTS:     4
TYPE OF CONTRAST:          CLASS                      NUMBER OF INDEPENDENT CONTRASTS:         3

001 STGRD<1  (1,M        ) 000                        GRADES IN STATE:  LESS THAN 1 GRADE, MISSING
002 STGRD1-2 (2          ) 100                        GRADES IN STATE:  1-2 GRADES
003 STGRD3-5 (3          ) 010                        GRADES IN STATE:  3-5 GRADES
004 STGRD6>  (4,5        ) 001                        GRADES IN STATE:  MORE THAN 5 GRADES

CONDITIONING VARIABLE ID:  BACK0032
DESCRIPTION:               DID YOU GO TO PRESCHOOL, NURSERY OR DAYCARE?
GRADES/ASSESSMENTS:        N04, S04
CONDITIONING VAR LABEL:    PRESCH
NAEP ID:                   B004201                    TOTAL NUMBER OF SPECIFIED CONTRASTS:     2
TYPE OF CONTRAST:          CLASS                      NUMBER OF INDEPENDENT CONTRASTS:         1

001 PRESCH-Y (2,3,M      ) 0                          PRESCHOOL:  NO, I DON'T KNOW, MISSING
002 PRESCH-N (1          ) 1                          PRESCHOOL:  YES

CONDITIONING VARIABLE ID:  BACK0033
DESCRIPTION:               HOW MANY TIMES HAVE YOU CHANGED SCHOOLS IN PAST TWO YEARS BECAUSE YOU MOVED?
GRADES/ASSESSMENTS:        N04, S04, N08, N12
CONDITIONING VAR LABEL:    SCH CHGS
NAEP ID:                   B007301                    TOTAL NUMBER OF SPECIFIED CONTRASTS:     4
TYPE OF CONTRAST:          CLASS                      NUMBER OF INDEPENDENT CONTRASTS:         3

001 SCHCHG-0 (1          ) 000                        SCHOOL CHANGES:  NONE
002 SCHCHG-1 (2          ) 100                        SCHOOL CHANGES:  ONE
003 SCHCHG-2 (3          ) 010                        SCHOOL CHANGES:  TWO
004 SCHCHG-3 (4,M        ) 001                        SCHOOL CHANGES:  THREE OR MORE, MISSING

CONDITIONING VARIABLE ID:  BACK0034
DESCRIPTION:               HOW OFTEN DO YOU DISCUSS THINGS STUDIED IN SCHOOL WITH SOMEONE AT HOME?
GRADES/ASSESSMENTS:        N04, S04, N08, N12
CONDITIONING VAR LABEL:    DISC@HOM
NAEP ID:                   B007401                    TOTAL NUMBER OF SPECIFIED CONTRASTS:     4
TYPE OF CONTRAST:          CLASS                      NUMBER OF INDEPENDENT CONTRASTS:         3
```

Table C-2 (continued)

Conditioning Variables Common to the Main Reading, U.S. History, and Geography Assessments

```
001 DIS@HOM1 (1          ) 000                              DISCUSS STUDIES AT HOME:  ALMOST EVERY DAY
002 DIS@HOM2 (2          ) 100                              DISCUSS STUDIES AT HOME:  ONCE OR TWICE A WEEK
003 DIS@HOM3 (3          ) 010                              DISCUSS STUDIES AT HOME:  ONCE OR TWICE A MONTH
004 DIS@HOM4 (4,M        ) 001                              DISCUSS STUDIES AT HOME:  NEVER OR HARDLY EVER,
                                                                                      MISSING

CONDITIONING VARIABLE ID:  BACK0035
DESCRIPTION:               ABOUT HOW MANY PAGES A DAY DO YOU HAVE TO READ FOR SCHOOL AND HOMEWORK?
GRADES/ASSESSMENTS:        N04, S04, N08, N12
CONDITIONING VAR LABEL:    PGSREAD1
NAEP ID:                   B001101                          TOTAL NUMBER OF SPECIFIED CONTRASTS:     2
TYPE OF CONTRAST:          CLASS                            NUMBER OF INDEPENDENT CONTRASTS:         1

001 PGS<6,?  (5,M        ) 0                                PAGES READ:  5 OR FEWER A DAY, MISSING
002 PGS>5    (1,2,3,4    ) 1                                PAGES READ:  6-10, 11-15, 16-20, 20 OR MORE

CONDITIONING VARIABLE ID:  BACK0036
DESCRIPTION:               ABOUT HOW MANY PAGES A DAY DO YOU HAVE TO READ FOR SCHOOL AND HOMEWORK?
GRADES/ASSESSMENTS:        N04, S04, N08, N12
CONDITIONING VAR LABEL:    PGSREAD2
NAEP ID:                   B001101                          TOTAL NUMBER OF SPECIFIED CONTRASTS:     2
TYPE OF CONTRAST:          CLASS                            NUMBER OF INDEPENDENT CONTRASTS:         1

001 PGS<11,? (4,5,M      ) 0                                PAGES READ:  6-10, 5 OR FEWER A DAY, MISSING
002 PGS>10   (1,2,3      ) 1                                PAGES READ:  11-15, 16-20, 20 OR MORE

CONDITIONING VARIABLE ID:  BACK0037
DESCRIPTION:               HOW OFTEN DO YOU USE A COMPUTER FOR SCHOOLWORK?
GRADES/ASSESSMENTS:        N04, S04, N08, N12
CONDITIONING VAR LABEL:    COMP@SCH
NAEP ID:                   B007501                          TOTAL NUMBER OF SPECIFIED CONTRASTS:     5
TYPE OF CONTRAST:          CLASS                            NUMBER OF INDEPENDENT CONTRASTS:         4

001 COMP-DAY (1          ) 0000                             USE COMPUTER AT SCHOOL:  ALMOST EVERY DAY
002 COMP-WK  (2          ) 1000                             USE COMPUTER AT SCHOOL:  ONCE OR TWICE A WEEK
003 COMP-MO  (3          ) 0100                             USE COMPUTER AT SCHOOL:  ONCE OR TWICE A MONTH
004 COMP-NEV (4          ) 0010                             USE COMPUTER AT SCHOOL:  NEVER OR HARDLY EVER
005 COMP-?   (M          ) 0001                             USE COMPUTER AT SCHOOL:  MISSING

CONDITIONING VARIABLE ID:  BACK0038
DESCRIPTION:               WHICH BEST DESCRIBES YOUR HIGH-SCHOOL PROGRAM?
GRADES/ASSESSMENTS:        N12
CONDITIONING VAR LABEL:    HS PROG
NAEP ID:                   B008501                          TOTAL NUMBER OF SPECIFIED CONTRASTS:     5
TYPE OF CONTRAST:          CLASS                            NUMBER OF INDEPENDENT CONTRASTS:         4

001 GENERAL  (1          ) 0000                             HIGH SCHOOL PROGRAM:  GENERAL
002 ACADEMIC (2          ) 1000                             HIGH SCHOOL PROGRAM:  ACADEMIC/COLLEGE PREP
003 VOC/TECH (3          ) 0100                             HIGH SCHOOL PROGRAM:  VOCATIONAL OR TECHNICAL
004 OTHERPGM (4          ) 0010                             HIGH SCHOOL PROGRAM:  OTHER
005 HS PGM-? (M          ) 0001                             HIGH SCHOOL PROGRAM:  MISSING

CONDITIONING VARIABLE ID:  BACK0039
DESCRIPTION:               WHAT DO YOU EXPECT TO BE YOUR MAIN ACTIVITY IN THE YEAR AFTER YOU LEAVE HIGH
                           SCHOOL?
GRADES/ASSESSMENTS:        N12
CONDITIONING VAR LABEL:    AFTER HS
NAEP ID:                   B008401                          TOTAL NUMBER OF SPECIFIED CONTRASTS:     7
TYPE OF CONTRAST:          CLASS                            NUMBER OF INDEPENDENT CONTRASTS:         6
```

Table C-2 (continued)

Conditioning Variables Common to the Main Reading, U.S. History, and Geography Assessments

```
001 WORKING  (1          ) 000000                AFTER HIGH SCHOOL:  WORKING FULL-TIME
002 VOC/BUSN (2          ) 100000                AFTER HIGH SCHOOL:  VOC/TECH/BUSINESS SCHOOL
003 2 YR COL (3          ) 010000                AFTER HIGH SCHOOL:  TWO-YEAR COLLEGE
004 4 YR COL (4          ) 001000                AFTER HIGH SCHOOL:  FOUR-YEAR COLLEGE
005 MILITARY (5          ) 000100                AFTER HIGH SCHOOL:  REGULAR MILITARY
006 OTHERACT (6          ) 000010                AFTER HIGH SCHOOL:  OTHER
007 AFTRHS-? (M          ) 000001                AFTER HIGH SCHOOL:  MISSING

CONDITIONING VARIABLE ID:  BACK0040
DESCRIPTION:               SEMESTERS ENGLISH/LITERATURE/WRITING (MISSING VS NON-MISSING)
GRADES/ASSESSMENTS:        N12
CONDITIONING VAR LABEL:    SEM ENG
NAEP ID:                   B007101               TOTAL NUMBER OF SPECIFIED CONTRASTS:     2
TYPE OF CONTRAST:          CLASS                 NUMBER OF INDEPENDENT CONTRASTS:         1

001 SEMENG-? (M          ) 0                     ENGLISH SEMESTERS:  MISSING
002 SEMENG-^ (1-9        ) 1                     ENGLISH SEMESTERS:  NOT MISSING

CONDITIONING VARIABLE ID:  BACK0041
DESCRIPTION:               NUMBER OF SEMESTERS ENGLISH/LITERATURE/WRITING (LINEAR)
GRADES/ASSESSMENTS:        N12
CONDITIONING VAR LABEL:    #ENG-LIN
NAEP ID:                   B007101               TOTAL NUMBER OF SPECIFIED CONTRASTS:     1
TYPE OF CONTRAST:          LINEAR                NUMBER OF INDEPENDENT CONTRASTS:         1

001 #ENG-LIN (1-9,M=0    )  0.0 +  1.0*X         NUMBER OF SEMESTERS ENGLISH/LITERATURE/WRITING
                                                 (LINEAR)

CONDITIONING VARIABLE ID:  BACK0042
DESCRIPTION:               SEMESTERS MATHEMATICS (MISSING VS NON-MISSING)
GRADES/ASSESSMENTS:        N12
CONDITIONING VAR LABEL:    SEM MAT
NAEP ID:                   B007102               TOTAL NUMBER OF SPECIFIED CONTRASTS:     2
TYPE OF CONTRAST:          CLASS                 NUMBER OF INDEPENDENT CONTRASTS:         1

001 SEMMAT-? (M          ) 0                     MATH SEMESTERS:  MISSING
002 SEMMAT-^ (1-9        ) 1                     MATH SEMESTERS:  NOT MISSING

CONDITIONING VARIABLE ID:  BACK0043
DESCRIPTION:               NUMBER OF SEMESTERS MATHEMATICS (LINEAR)
GRADES/ASSESSMENTS:        N12
CONDITIONING VAR LABEL:    #MAT-LIN
NAEP ID:                   B007102               TOTAL NUMBER OF SPECIFIED CONTRASTS:     1
TYPE OF CONTRAST:          LINEAR                NUMBER OF INDEPENDENT CONTRASTS:         1

001 #MAT-LIN (1-9,M=0    )  0.0 +  1.0*X         NUMBER OF SEMESTERS MATHEMATICS (LINEAR)

CONDITIONING VARIABLE ID:  BACK0044
DESCRIPTION:               SEMESTERS SCIENCE (MISSING VS NON-MISSING)
GRADES/ASSESSMENTS:        N12
CONDITIONING VAR LABEL:    SEM SCI
NAEP ID:                   B007103               TOTAL NUMBER OF SPECIFIED CONTRASTS:     2
TYPE OF CONTRAST:          CLASS                 NUMBER OF INDEPENDENT CONTRASTS:         1

001 SEMSCI-? (M          ) 0                     SCIENCE SEMESTERS:  MISSING
002 SEMSCI-^ (1-9        ) 1                     SCIENCE SEMESTERS:  NOT MISSING

CONDITIONING VARIABLE ID:  BACK0045
DESCRIPTION:               NUMBER OF SEMESTERS SCIENCE (LINEAR)
GRADES/ASSESSMENTS:        N12
CONDITIONING VAR LABEL:    #SCI-LIN
NAEP ID:                   B007103               TOTAL NUMBER OF SPECIFIED CONTRASTS:     1
TYPE OF CONTRAST:          LINEAR                NUMBER OF INDEPENDENT CONTRASTS:         1
```

Table C-2 (continued)

Conditioning Variables Common to the Main Reading, U.S. History, and Geography Assessments

```
001 #SCI-LIN (1-9,M=0    )  0.0 +  1.0*X                NUMBER OF SEMESTERS SCIENCE (LINEAR)

CONDITIONING VARIABLE ID:  BACK0046
DESCRIPTION:               SEMESTERS HISTORY/SOCIAL STUDIES/GEOGRAPHY (MISSING VS NON-MISSING)
GRADES/ASSESSMENTS:        N12
CONDITIONING VAR LABEL:    SEM HIS
NAEP ID:                   B007104                      TOTAL NUMBER OF SPECIFIED CONTRASTS:     2
TYPE OF CONTRAST:          CLASS                        NUMBER OF INDEPENDENT CONTRASTS:         1

001 SEMHIS-? (M          ) 0                            HISTORY SEMESTERS:  MISSING
002 SEMHIS-^ (1-9        ) 1                            HISTORY SEMESTERS:  NOT MISSING

CONDITIONING VARIABLE ID:  BACK0047
DESCRIPTION:               NUMBER OF SEMESTERS HISTORY/SOCIAL STUDIES/GEOGRAPHY (LINEAR)
GRADES/ASSESSMENTS:        N12
CONDITIONING VAR LABEL:    #HIS-LIN
NAEP ID:                   B007104                      TOTAL NUMBER OF SPECIFIED CONTRASTS:     1
TYPE OF CONTRAST:          LINEAR                       NUMBER OF INDEPENDENT CONTRASTS:         1

001 #HIS-LIN (1-9,M=0    )  0.0 +  1.0*X                NUMBER OF SEMESTERS HISTORY/SOCIAL
                                                STUDIES/GEOGRAPHY (LINEAR)

CONDITIONING VARIABLE ID:  BACK0048
DESCRIPTION:               SEMESTERS FOREIGN LANGUAGES (MISSING VS NON-MISSING)
GRADES/ASSESSMENTS:        N12
CONDITIONING VAR LABEL:    SEM FLG
NAEP ID:                   B007105                      TOTAL NUMBER OF SPECIFIED CONTRASTS:     2
TYPE OF CONTRAST:          CLASS                        NUMBER OF INDEPENDENT CONTRASTS:         1

001 SEMFLG-? (M          ) 0                            FOREIGN LANGUAGE SEMESTERS:  MISSING
002 SEMFLG-^ (1-9        ) 1                            FOREIGN LANGUAGE SEMESTERS:  NOT MISSING

CONDITIONING VARIABLE ID:  BACK0049
DESCRIPTION:               NUMBER OF SEMESTERS FOREIGN LANGUAGES (LINEAR)
GRADES/ASSESSMENTS:        N12
CONDITIONING VAR LABEL:    #FLG-LIN
NAEP ID:                   B007105                      TOTAL NUMBER OF SPECIFIED CONTRASTS:     1
TYPE OF CONTRAST:          LINEAR                       NUMBER OF INDEPENDENT CONTRASTS:         1

001 #FLG-LIN (1-9,M=0    )  0.0 +  1.0*X                NUMBER OF SEMESTERS FOREIGN LANGUAGES (LINEAR)

CONDITIONING VARIABLE ID:  BACK0050
DESCRIPTION:               SEMESTERS VOCATIONAL/TECHNICAL/BUSINESS EDUCATION (MISSING VS NON-MISSING)
GRADES/ASSESSMENTS:        N12
CONDITIONING VAR LABEL:    SEM VOC
NAEP ID:                   B007106                      TOTAL NUMBER OF SPECIFIED CONTRASTS:     2
TYPE OF CONTRAST:          CLASS                        NUMBER OF INDEPENDENT CONTRASTS:         1

001 SEMVOC-? (M          ) 0                            VOC/TECH SEMESTERS:  MISSING
002 SEMVOC-^ (1-9        ) 1                            VOC/TECH SEMESTERS:  NOT MISSING

CONDITIONING VARIABLE ID:  BACK0051
DESCRIPTION:               NUMBER OF SEMESTERS VOCATIONAL/TECHNICAL/BUSINESS EDUCATION (LINEAR)
GRADES/ASSESSMENTS:        N12
CONDITIONING VAR LABEL:    #VOC-LIN
NAEP ID:                   B007106                      TOTAL NUMBER OF SPECIFIED CONTRASTS:     1
TYPE OF CONTRAST:          LINEAR                       NUMBER OF INDEPENDENT CONTRASTS:         1

001 #VOC-LIN (1-9,M=0    )  0.0 +  1.0*X                NUMBER OF SEMESTERS VOCATIONAL/TECHNICAL/BUSINESS
                                                        EDUCATION (LINEAR)

CONDITIONING VARIABLE ID:  BACK0052
DESCRIPTION:               SEMESTERS ART/MUSIC (MISSING VS NON-MISSING)
GRADES/ASSESSMENTS:        N12
```

Table C-2 (continued)

Conditioning Variables Common to the Main Reading, U.S. History, and Geography Assessments

```
CONDITIONING VAR LABEL:     SEM ART
NAEP ID:                    B007107                   TOTAL NUMBER OF SPECIFIED CONTRASTS:    2
TYPE OF CONTRAST:           CLASS                     NUMBER OF INDEPENDENT CONTRASTS:        1

001 SEMART-? (M           ) 0                         ART/MUSIC SEMESTERS:  MISSING
002 SEMART-^ (1-9         ) 1                         ART/MUSIC SEMESTERS:  NOT MISSING

CONDITIONING VARIABLE ID:  BACK0053
DESCRIPTION:               NUMBER OF SEMESTERS ART/MUSIC (LINEAR)
GRADES/ASSESSMENTS:        N12
CONDITIONING VAR LABEL:    #ART-LIN
NAEP ID:                   B007107                    TOTAL NUMBER OF SPECIFIED CONTRASTS:    1
TYPE OF CONTRAST:          LINEAR                     NUMBER OF INDEPENDENT CONTRASTS:        1

001 #ART-LIN (1-9,M=0     )  0.0 +  1.0*X             NUMBER OF SEMESTERS ART/MUSIC (LINEAR)

CONDITIONING VARIABLE ID:  BACK0054
DESCRIPTION:               INTERACTION:  GENDER BY RACE/ETHNICITY
GRADES/ASSESSMENTS:        N04, S04, N08, N12
CONDITIONING VAR LABEL:    GEND/RAC
NAEP ID:                   N/A                        TOTAL NUMBER OF SPECIFIED CONTRASTS:   10
TYPE OF CONTRAST:          INTERACTION                NUMBER OF INDEPENDENT CONTRASTS:        4

001 G/R 11    (11         ) 01010101                  GEND/RAC INTACT: 1. MALE      1. WHI/AI/O
002 G/R 12    (12         ) -1000000                  GEND/RAC INTACT: 1. MALE      2. BLACK
003 G/R 13    (13         ) 00-10000                  GEND/RAC INTACT: 1. MALE      3. HISPANIC
004 G/R 14    (14         ) 0000-100                  GEND/RAC INTACT: 1. MALE      4. ASIAN
005 G/R 15    (15         ) 000000-1                  GEND/RAC INTACT: 1. MALE      5. PAC ISLD
006 G/R 21    (21         ) -1-1-1-1                  GEND/RAC INTACT: 2. FEMALE    1. WHI/AI/O
007 G/R 22    (22         ) 01000000                  GEND/RAC INTACT: 2. FEMALE    2. BLACK
008 G/R 23    (23         ) 00010000                  GEND/RAC INTACT: 2. FEMALE    3. HISPANIC
009 G/R 24    (24         ) 00000100                  GEND/RAC INTACT: 2. FEMALE    4. ASIAN
010 G/R 25    (25         ) 00000001                  GEND/RAC INTACT: 2. FEMALE    5. PAC ISLD

CONDITIONING VARIABLE ID:  BACK0055
DESCRIPTION:               INTERACTION:  GENDER BY TYPE OF LOCALE (5 CATEGORIES)
GRADES/ASSESSMENTS:        N04, S04, N08, N12
CONDITIONING VAR LABEL:    GEND/TOL
NAEP ID:                   N/A                        TOTAL NUMBER OF SPECIFIED CONTRASTS:   10
TYPE OF CONTRAST:          INTERACTION                NUMBER OF INDEPENDENT CONTRASTS:        4

001 G/T 11    (11         ) 01010101                  GEND/TOL INTACT: 1. MALE      1. BIG CTY5
002 G/T 12    (12         ) -1000000                  GEND/TOL INTACT: 1. MALE      2. MID CTY5
003 G/T 13    (13         ) 00-10000                  GEND/TOL INTACT: 1. MALE      3. FR/BTWN5
004 G/T 14    (14         ) 0000-100                  GEND/TOL INTACT: 1. MALE      4. SML TWN5
005 G/T 15    (15         ) 000000-1                  GEND/TOL INTACT: 1. MALE      5. RURAL5
006 G/T 21    (21         ) -1-1-1-1                  GEND/TOL INTACT: 2. FEMALE    1. BIG CTY5
007 G/T 22    (22         ) 01000000                  GEND/TOL INTACT: 2. FEMALE    2. MID CTY5
008 G/T 23    (23         ) 00010000                  GEND/TOL INTACT: 2. FEMALE    3. FR/BTWN5
009 G/T 24    (24         ) 00000100                  GEND/TOL INTACT: 2. FEMALE    4. SML TWN5
010 G/T 25    (25         ) 00000001                  GEND/TOL INTACT: 2. FEMALE    5. RURAL5

CONDITIONING VARIABLE ID:  BACK0056
DESCRIPTION:               INTERACTION:  GENDER BY PARENTS' EDUCATION
GRADES/ASSESSMENTS:        N04, S04, N08, N12
CONDITIONING VAR LABEL:    GEND/PAR
NAEP ID:                   N/A                        TOTAL NUMBER OF SPECIFIED CONTRASTS:   10
TYPE OF CONTRAST:          INTERACTION                NUMBER OF INDEPENDENT CONTRASTS:        4
```

Table C-2 (continued)

Conditioning Variables Common to the Main Reading, U.S. History, and Geography Assessments

```
001 G/P 11     (11           ) 01010101          GEND/PAR INTACT: 1. MALE     1. < HS
002 G/P 12     (12           ) -1000000          GEND/PAR INTACT: 1. MALE     2. HS GRAD
003 G/P 13     (13           ) 00-10000          GEND/PAR INTACT: 1. MALE     3. POST HS
004 G/P 14     (14           ) 0000-100          GEND/PAR INTACT: 1. MALE     4. COL GRAD
005 G/P 15     (15           ) 000000-1          GEND/PAR INTACT: 1. MALE     5. PARED-?
006 G/P 21     (21           ) -1-1-1-1          GEND/PAR INTACT: 2. FEMALE   1. < HS
007 G/P 22     (22           ) 01000000          GEND/PAR INTACT: 2. FEMALE   2. HS GRAD
008 G/P 23     (23           ) 00010000          GEND/PAR INTACT: 2. FEMALE   3. POST HS
009 G/P 24     (24           ) 00000100          GEND/PAR INTACT: 2. FEMALE   4. COL GRAD
010 G/P 25     (25           ) 00000001          GEND/PAR INTACT: 2. FEMALE   5. PARED-?

CONDITIONING VARIABLE ID:    BACK0057
DESCRIPTION:                 INTERACTION:  GENDER BY SCHOOL TYPE
GRADES/ASSESSMENTS:          N04, S04, N08, N12
CONDITIONING VAR LABEL:      GEND/SCH
NAEP ID:                     N/A                      TOTAL NUMBER OF SPECIFIED CONTRASTS:     6
TYPE OF CONTRAST:            INTERACTION              NUMBER OF INDEPENDENT CONTRASTS:         2

001 G/S 11     (11           ) 0101              GEND/SCH INTACT: 1. MALE     1. PUBLIC
002 G/S 12     (12           ) -100              GEND/SCH INTACT: 1. MALE     2. PRIVATE
003 G/S 13     (13           ) 00-1              GEND/SCH INTACT: 1. MALE     3. CATHOLIC
004 G/S 21     (21           ) -1-1              GEND/SCH INTACT: 2. FEMALE   1. PUBLIC
005 G/S 22     (22           ) 0100              GEND/SCH INTACT: 2. FEMALE   2. PRIVATE
006 G/S 23     (23           ) 0001              GEND/SCH INTACT: 2. FEMALE   3. CATHOLIC

CONDITIONING VARIABLE ID:    BACK0058
DESCRIPTION:                 INTERACTION:  RACE/ETHNICITY BY TYPE OF LOCALE (5 CATEGORIES)
GRADES/ASSESSMENTS:          N04, S04, N08, N12
CONDITIONING VAR LABEL:      RACE/TOL
NAEP ID:                     N/A                      TOTAL NUMBER OF SPECIFIED CONTRASTS:    25
TYPE OF CONTRAST:            INTERACTION              NUMBER OF INDEPENDENT CONTRASTS:        16

001 R/T 11     (11           ) 01010101010101010101010101010101 RACE/TOL INTACT: 1. WHI/AI/O 1. BIG CTY5
002 R/T 12     (12           ) -1000000-1000000-1000000-1000000 RACE/TOL INTACT: 1. WHI/AI/O 2. MID CTY5
003 R/T 13     (13           ) 00-1000000-1000000-1000000-10000 RACE/TOL INTACT: 1. WHI/AI/O 3. FR/BTWN5
004 R/T 14     (14           ) 0000-1000000-1000000-1000000-100 RACE/TOL INTACT: 1. WHI/AI/O 4. SML TWN5
005 R/T 15     (15           ) 000000-1000000-1000000-1000000-1 RACE/TOL INTACT: 1. WHI/AI/O 5. RURAL5
006 R/T 21     (21           ) -1-1-1-1000000000000000000000000 RACE/TOL INTACT: 2. BLACK    1. BIG CTY5
007 R/T 22     (22           ) 01000000000000000000000000000000 RACE/TOL INTACT: 2. BLACK    2. MID CTY5
008 R/T 23     (23           ) 00010000000000000000000000000000 RACE/TOL INTACT: 2. BLACK    3. FR/BTWN5
009 R/T 24     (24           ) 00000100000000000000000000000000 RACE/TOL INTACT: 2. BLACK    4. SML TWN5
010 R/T 25     (25           ) 00000001000000000000000000000000 RACE/TOL INTACT: 2. BLACK    5. RURAL5
011 R/T 31     (31           ) 00000000-1-1-1-10000000000000000 RACE/TOL INTACT: 3. HISPANIC 1. BIG CTY5
012 R/T 32     (32           ) 00000000010000000000000000000000 RACE/TOL INTACT: 3. HISPANIC 2. MID CTY5
013 R/T 33     (33           ) 00000000000100000000000000000000 RACE/TOL INTACT: 3. HISPANIC 3. FR/BTWN5
014 R/T 34     (34           ) 00000000000001000000000000000000 RACE/TOL INTACT: 3. HISPANIC 4. SML TWN5
015 R/T 35     (35           ) 00000000000000010000000000000000 RACE/TOL INTACT: 3. HISPANIC 5. RURAL5
016 R/T 41     (41           ) 0000000000000000-1-1-1-100000000 RACE/TOL INTACT: 4. ASIAN    1. BIG CTY5
017 R/T 42     (42           ) 00000000000000000100000000000000 RACE/TOL INTACT: 4. ASIAN    2. MID CTY5
018 R/T 43     (43           ) 00000000000000000001000000000000 RACE/TOL INTACT: 4. ASIAN    3. FR/BTWN5
019 R/T 44     (44           ) 00000000000000000000010000000000 RACE/TOL INTACT: 4. ASIAN    4. SML TWN5
020 R/T 45     (45           ) 00000000000000000000000100000000 RACE/TOL INTACT: 4. ASIAN    5. RURAL5
021 R/T 51     (51           ) 000000000000000000000000-1-1-1-1 RACE/TOL INTACT: 5. PAC ISLD 1. BIG CTY5
022 R/T 52     (52           ) 00000000000000000000000001000000 RACE/TOL INTACT: 5. PAC ISLD 2. MID CTY5
023 R/T 53     (53           ) 00000000000000000000000000010000 RACE/TOL INTACT: 5. PAC ISLD 3. FR/BTWN5
024 R/T 54     (54           ) 00000000000000000000000000000100 RACE/TOL INTACT: 5. PAC ISLD 4. SML TWN5
025 R/T 55     (55           ) 00000000000000000000000000000001 RACE/TOL INTACT: 5. PAC ISLD 5. RURAL5

CONDITIONING VARIABLE ID:    BACK0059
DESCRIPTION:                 INTERACTION:  RACE/ETHNICITY BY PARENTS' EDUCATION
GRADES/ASSESSMENTS:          N04, S04, N08, N12
CONDITIONING VAR LABEL:      RACE/PAR
NAEP ID:                     N/A                      TOTAL NUMBER OF SPECIFIED CONTRASTS:    25
TYPE OF CONTRAST:            INTERACTION              NUMBER OF INDEPENDENT CONTRASTS:        16
```

Table C-2 (continued)

Conditioning Variables Common to the Main Reading, U.S. History, and Geography Assessments

```
001 R/P 11   (11         ) 01010101010101010101010101010101 RACE/PAR INTACT: 1. WHI/AI/O 1. < HS
002 R/P 12   (12         ) -1000000-1000000-1000000-1000000 RACE/PAR INTACT: 1. WHI/AI/O 2. HS GRAD
003 R/P 13   (13         ) 00-1000000-1000000-1000000-10000 RACE/PAR INTACT: 1. WHI/AI/O 3. POST HS
004 R/P 14   (14         ) 0000-1000000-1000000-1000000-100 RACE/PAR INTACT: 1. WHI/AI/O 4. COL GRAD
005 R/P 15   (15         ) 000000-1000000-1000000-1000000-1 RACE/PAR INTACT: 1. WHI/AI/O 5. PARED-?
006 R/P 21   (21         ) -1-1-1-1000000000000000000000000 RACE/PAR INTACT: 2. BLACK    1. < HS
007 R/P 22   (22         ) 01000000000000000000000000000000 RACE/PAR INTACT: 2. BLACK    2. HS GRAD
008 R/P 23   (23         ) 00010000000000000000000000000000 RACE/PAR INTACT: 2. BLACK    3. POST HS
009 R/P 24   (24         ) 00000100000000000000000000000000 RACE/PAR INTACT: 2. BLACK    4. COL GRAD
010 R/P 25   (25         ) 00000001000000000000000000000000 RACE/PAR INTACT: 2. BLACK    5. PARED-?
011 R/P 31   (31         ) 00000000-1-1-1-10000000000000000 RACE/PAR INTACT: 3. HISPANIC 1. < HS
012 R/P 32   (32         ) 00000000010000000000000000000000 RACE/PAR INTACT: 3. HISPANIC 2. HS GRAD
013 R/P 33   (33         ) 00000000000100000000000000000000 RACE/PAR INTACT: 3. HISPANIC 3. POST HS
014 R/P 34   (34         ) 00000000000001000000000000000000 RACE/PAR INTACT: 3. HISPANIC 4. COL GRAD
015 R/P 35   (35         ) 00000000000000010000000000000000 RACE/PAR INTACT: 3. HISPANIC 5. PARED-?
016 R/P 41   (41         ) 0000000000000000-1-1-1-100000000 RACE/PAR INTACT: 4. ASIAN    1. < HS
017 R/P 42   (42         ) 00000000000000000100000000000000 RACE/PAR INTACT: 4. ASIAN    2. HS GRAD
018 R/P 43   (43         ) 00000000000000000001000000000000 RACE/PAR INTACT: 4. ASIAN    3. POST HS
019 R/P 44   (44         ) 00000000000000000000010000000000 RACE/PAR INTACT: 4. ASIAN    4. COL GRAD
020 R/P 45   (45         ) 00000000000000000000000100000000 RACE/PAR INTACT: 4. ASIAN    5. PARED-?
021 R/P 51   (51         ) 000000000000000000000000-1-1-1-1 RACE/PAR INTACT: 5. PAC ISLD 1. < HS
022 R/P 52   (52         ) 00000000000000000000000001000000 RACE/PAR INTACT: 5. PAC ISLD 2. HS GRAD
023 R/P 53   (53         ) 00000000000000000000000000010000 RACE/PAR INTACT: 5. PAC ISLD 3. POST HS
024 R/P 54   (54         ) 00000000000000000000000000000100 RACE/PAR INTACT: 5. PAC ISLD 4. COL GRAD
025 R/P 55   (55         ) 00000000000000000000000000000001 RACE/PAR INTACT: 5. PAC ISLD 5. PARED-?

CONDITIONING VARIABLE ID:  BACK0060
DESCRIPTION:               INTERACTION:  RACE/ETHNICITY BY SCHOOL TYPE
GRADES/ASSESSMENTS:        N04, S04, N08, N12
CONDITIONING VAR LABEL:    RACE/SCH
NAEP ID:                   N/A                              TOTAL NUMBER OF SPECIFIED CONTRASTS:    15
TYPE OF CONTRAST:          INTERACTION                      NUMBER OF INDEPENDENT CONTRASTS:         8

001 R/S 11   (11         ) 0101010101010101                 RACE/SCH INTACT: 1. WHI/AI/O 1. PUBLIC
002 R/S 12   (12         ) -100-100-100-100                 RACE/SCH INTACT: 1. WHI/AI/O 2. PRIVATE
003 R/S 13   (13         ) 00-100-100-100-1                 RACE/SCH INTACT: 1. WHI/AI/O 3. CATHOLIC
004 R/S 21   (21         ) -1-1000000000000                 RACE/SCH INTACT: 2. BLACK    1. PUBLIC
005 R/S 22   (22         ) 0100000000000000                 RACE/SCH INTACT: 2. BLACK    2. PRIVATE
006 R/S 23   (23         ) 0001000000000000                 RACE/SCH INTACT: 2. BLACK    3. CATHOLIC
007 R/S 31   (31         ) 0000-1-100000000                 RACE/SCH INTACT: 3. HISPANIC 1. PUBLIC
008 R/S 32   (32         ) 0000010000000000                 RACE/SCH INTACT: 3. HISPANIC 2. PRIVATE
009 R/S 33   (33         ) 0000000100000000                 RACE/SCH INTACT: 3. HISPANIC 3. CATHOLIC
010 R/S 41   (41         ) 00000000-1-10000                 RACE/SCH INTACT: 4. ASIAN    1. PUBLIC
011 R/S 42   (42         ) 0000000001000000                 RACE/SCH INTACT: 4. ASIAN    2. PRIVATE
012 R/S 43   (43         ) 0000000000010000                 RACE/SCH INTACT: 4. ASIAN    3. CATHOLIC
013 R/S 51   (51         ) 000000000000-1-1                 RACE/SCH INTACT: 5. PAC ISLD 1. PUBLIC
014 R/S 52   (52         ) 0000000000000100                 RACE/SCH INTACT: 5. PAC ISLD 2. PRIVATE
015 R/S 53   (53         ) 0000000000000001                 RACE/SCH INTACT: 5. PAC ISLD 3. CATHOLIC

CONDITIONING VARIABLE ID:  BACK0061
DESCRIPTION:               INTERACTION:  TYPE OF LOCALE (5 CATEGORIES) BY PARENT'S EDUCATION
GRADES/ASSESSMENTS:        N04, S04, N08, N12
CONDITIONING VAR LABEL:    TOL5/PAR
NAEP ID:                   N/A                              TOTAL NUMBER OF SPECIFIED CONTRASTS:    25
TYPE OF CONTRAST:          INTERACTION                      NUMBER OF INDEPENDENT CONTRASTS:        16
```

Table C-2 (continued)

Conditioning Variables Common to the Main Reading, U.S. History, and Geography Assessments

```
001 T/P 11   (11          ) 01010101010101010101010101010101 TOL5/PAR INTACT: 1. BIG CTY5 1. < HS
002 T/P 12   (12          ) -1000000-1000000-1000000-1000000 TOL5/PAR INTACT: 1. BIG CTY5 2. HS GRAD
003 T/P 13   (13          ) 00-1000000-1000000-1000000-10000 TOL5/PAR INTACT: 1. BIG CTY5 3. POST HS
004 T/P 14   (14          ) 0000-1000000-1000000-1000000-100 TOL5/PAR INTACT: 1. BIG CTY5 4. COL GRAD
005 T/P 15   (15          ) 000000-1000000-1000000-1000000-1 TOL5/PAR INTACT: 1. BIG CTY5 5. PARED-?
006 T/P 21   (21          ) -1-1-1-1000000000000000000000000 TOL5/PAR INTACT: 2. MID CTY5 1. < HS
007 T/P 22   (22          ) 01000000000000000000000000000000 TOL5/PAR INTACT: 2. MID CTY5 2. HS GRAD
008 T/P 23   (23          ) 00010000000000000000000000000000 TOL5/PAR INTACT: 2. MID CTY5 3. POST HS
009 T/P 24   (24          ) 00000100000000000000000000000000 TOL5/PAR INTACT: 2. MID CTY5 4. COL GRAD
010 T/P 25   (25          ) 00000001000000000000000000000000 TOL5/PAR INTACT: 2. MID CTY5 5. PARED-?
011 T/P 31   (31          ) 00000000-1-1-1-10000000000000000 TOL5/PAR INTACT: 3. FR/BTWN5 1. < HS
012 T/P 32   (32          ) 00000000010000000000000000000000 TOL5/PAR INTACT: 3. FR/BTWN5 2. HS GRAD
013 T/P 33   (33          ) 00000000000100000000000000000000 TOL5/PAR INTACT: 3. FR/BTWN5 3. POST HS
014 T/P 34   (34          ) 00000000000001000000000000000000 TOL5/PAR INTACT: 3. FR/BTWN5 4. COL GRAD
015 T/P 35   (35          ) 00000000000000010000000000000000 TOL5/PAR INTACT: 3. FR/BTWN5 5. PARED-?
016 T/P 41   (41          ) 0000000000000000-1-1-1-100000000 TOL5/PAR INTACT: 4. SML TWN5 1. < HS
017 T/P 42   (42          ) 00000000000000000100000000000000 TOL5/PAR INTACT: 4. SML TWN5 2. HS GRAD
018 T/P 43   (43          ) 00000000000000000001000000000000 TOL5/PAR INTACT: 4. SML TWN5 3. POST HS
019 T/P 44   (44          ) 00000000000000000000010000000000 TOL5/PAR INTACT: 4. SML TWN5 4. COL GRAD
020 T/P 45   (45          ) 00000000000000000000000100000000 TOL5/PAR INTACT: 4. SML TWN5 5. PARED-?
021 T/P 51   (51          ) 000000000000000000000000-1-1-1-1 TOL5/PAR INTACT: 5. RURAL5   1. < HS
022 T/P 52   (52          ) 00000000000000000000000001000000 TOL5/PAR INTACT: 5. RURAL5   2. HS GRAD
023 T/P 53   (53          ) 00000000000000000000000000010000 TOL5/PAR INTACT: 5. RURAL5   3. POST HS
024 T/P 54   (54          ) 00000000000000000000000000000100 TOL5/PAR INTACT: 5. RURAL5   4. COL GRAD
025 T/P 55   (55          ) 00000000000000000000000000000001 TOL5/PAR INTACT: 5. RURAL5   5. PARED-?

CONDITIONING VARIABLE ID:  BACK0062
DESCRIPTION:               INTERACTION:  TYPE OF LOCALE (5 CATEGORIES) BY SCHOOL TYPE
GRADES/ASSESSMENTS:        N04, S04, N08, N12
CONDITIONING VAR LABEL:    TOL5/SCH
NAEP ID:                   N/A                              TOTAL NUMBER OF SPECIFIED CONTRASTS:    15
TYPE OF CONTRAST:          INTERACTION                      NUMBER OF INDEPENDENT CONTRASTS:         8

001 T/S 11   (11          ) 0101010101010101                 TOL5/SCH INTACT: 1. BIG CTY5 1. PUBLIC
002 T/S 12   (12          ) -100-100-100-100                 TOL5/SCH INTACT: 1. BIG CTY5 2. PRIVATE
003 T/S 13   (13          ) 00-100-100-100-1                 TOL5/SCH INTACT: 1. BIG CTY5 3. CATHOLIC
004 T/S 21   (21          ) -1-1000000000000                 TOL5/SCH INTACT: 2. MID CTY5 1. PUBLIC
005 T/S 22   (22          ) 0100000000000000                 TOL5/SCH INTACT: 2. MID CTY5 2. PRIVATE
006 T/S 23   (23          ) 0001000000000000                 TOL5/SCH INTACT: 2. MID CTY5 3. CATHOLIC
007 T/S 31   (31          ) 0000-1-100000000                 TOL5/SCH INTACT: 3. FR/BTWN5 1. PUBLIC
008 T/S 32   (32          ) 0000010000000000                 TOL5/SCH INTACT: 3. FR/BTWN5 2. PRIVATE
009 T/S 33   (33          ) 0000000100000000                 TOL5/SCH INTACT: 3. FR/BTWN5 3. CATHOLIC
010 T/S 41   (41          ) 00000000-1-10000                 TOL5/SCH INTACT: 4. SML TWN5 1. PUBLIC
011 T/S 42   (42          ) 0000000001000000                 TOL5/SCH INTACT: 4. SML TWN5 2. PRIVATE
012 T/S 43   (43          ) 0000000000010000                 TOL5/SCH INTACT: 4. SML TWN5 3. CATHOLIC
013 T/S 51   (51          ) 000000000000-1-1                 TOL5/SCH INTACT: 5. RURAL5   1. PUBLIC
014 T/S 52   (52          ) 0000000000000100                 TOL5/SCH INTACT: 5. RURAL5   2. PRIVATE
015 T/S 53   (53          ) 0000000000000001                 TOL5/SCH INTACT: 5. RURAL5   3. CATHOLIC

CONDITIONING VARIABLE ID:  BACK0063
DESCRIPTION:               INTERACTION:  PARENTS' EDUCATION BY SCHOOL TYPE
GRADES/ASSESSMENTS:        N04, S04, N08, N12
CONDITIONING VAR LABEL:    PARE/SCH
NAEP ID:                   N/A                              TOTAL NUMBER OF SPECIFIED CONTRASTS:    15
TYPE OF CONTRAST:          INTERACTION                      NUMBER OF INDEPENDENT CONTRASTS:         8

001 P/S 11   (11          ) 0101010101010101                 PARE/SCH INTACT: 1. < HS      1. PUBLIC
002 P/S 12   (12          ) -100-100-100-100                 PARE/SCH INTACT: 1. < HS      2. PRIVATE
003 P/S 13   (13          ) 00-100-100-100-1                 PARE/SCH INTACT: 1. < HS      3. CATHOLIC
004 P/S 21   (21          ) -1-1000000000000                 PARE/SCH INTACT: 2. HS GRAD   1. PUBLIC
005 P/S 22   (22          ) 0100000000000000                 PARE/SCH INTACT: 2. HS GRAD   2. PRIVATE
006 P/S 23   (23          ) 0001000000000000                 PARE/SCH INTACT: 2. HS GRAD   3. CATHOLIC
007 P/S 31   (31          ) 0000-1-100000000                 PARE/SCH INTACT: 3. POST HS   1. PUBLIC
008 P/S 32   (32          ) 0000010000000000                 PARE/SCH INTACT: 3. POST HS   2. PRIVATE
```

Table C-2 (continued)

Conditioning Variables Common to the Main Reading, U.S. History, and Geography Assessments

```
009 P/S 33   (33          ) 0000000100000000          PARE/SCH INTACT: 3. POST HS  3. CATHOLIC
010 P/S 41   (41          ) 00000000-1-10000          PARE/SCH INTACT: 4. COL GRAD 1. PUBLIC
011 P/S 42   (42          ) 0000000001000000          PARE/SCH INTACT: 4. COL GRAD 2. PRIVATE
012 P/S 43   (43          ) 0000000000010000          PARE/SCH INTACT: 4. COL GRAD 3. CATHOLIC
013 P/S 51   (51          ) 000000000000-1-1          PARE/SCH INTACT: 5. PARED-?  1. PUBLIC
014 P/S 52   (52          ) 0000000000000100          PARE/SCH INTACT: 5. PARED-?  2. PRIVATE
015 P/S 53   (53          ) 0000000000000001          PARE/SCH INTACT: 5. PARED-?  3. CATHOLIC

CONDITIONING VARIABLE ID:  SCHL0001
DESCRIPTION:               SCHOOL LEVEL AVERAGE READING NORMIT (MISSING VS NON-MISSING)
GRADES/ASSESSMENTS:        N04, S04, N08, N12
CONDITIONING VAR LABEL:    SCH NORM
NAEP ID:                   SCHNORM                    TOTAL NUMBER OF SPECIFIED CONTRASTS:     2
TYPE OF CONTRAST:          CLASS                      NUMBER OF INDEPENDENT CONTRASTS:         1

001 SCHNRM-? (M          ) 0                          SCHOOL LEVEL AVERAGE READING NORMIT MISSING
002 SCHNRM-Y (@          ) 1                          SCHOOL LEVEL AVERAGE READING NORMIT NOT-MISSING

CONDITIONING VARIABLE ID:  SCHL0002
DESCRIPTION:               SCHOOL LEVEL AVERAGE READING NORMIT
GRADES/ASSESSMENTS:        N04, S04, N08, N12
CONDITIONING VAR LABEL:    SNRM-LIN
NAEP ID:                   SCHNORM                    TOTAL NUMBER OF SPECIFIED CONTRASTS:     2
TYPE OF CONTRAST:          SCALE                      NUMBER OF INDEPENDENT CONTRASTS:         1

001 SNRM-LIN (#          ) (F8.4)                     SCHOOL LEVEL AVERAGE READING NORMIT MEAN
002 SNRM-LIN (M          ) 0                          SCHOOL LEVEL AVERAGE READING NORMIT MISSING

CONDITIONING VARIABLE ID:  SCHL0003
DESCRIPTION:               HOW IS 4TH GRADE ORGANIZED AT YOUR SCHOOL?
GRADES/ASSESSMENTS:        N04, S04
CONDITIONING VAR LABEL:    SCH ORG4
NAEP ID:                   C030900                    TOTAL NUMBER OF SPECIFIED CONTRASTS:     4
TYPE OF CONTRAST:          CLASS                      NUMBER OF INDEPENDENT CONTRASTS:         3

001 SELFCONT (1          ) 000                        4TH GRADE ORGANIZATION:  SELF CONTAINED
002 DEPTLIZD (2          ) 100                        4TH GRADE ORGANIZATION:  DEPARTMENTALIZED
003 REGROUPD (3          ) 010                        4TH GRADE ORGANIZATION:  REGROUPED
004 SCH4ORG? (M          ) 001                        4TH GRADE ORGANIZATION:  MISSING

CONDITIONING VARIABLE ID:  SCHL0004
DESCRIPTION:               HOW IS 8TH GRADE ORGANIZED AT YOUR SCHOOL?
GRADES/ASSESSMENTS:        N08
CONDITIONING VAR LABEL:    SCH ORG8
NAEP ID:                   C034200                    TOTAL NUMBER OF SPECIFIED CONTRASTS:     4
TYPE OF CONTRAST:          CLASS                      NUMBER OF INDEPENDENT CONTRASTS:         3

001 SELFCONT (1          ) 000                        8TH GRADE ORGANIZATION:  SELF CONTAINED
002 SEMIDEPT (2          ) 100                        8TH GRADE ORGANIZATION:  SEMI-DEPARTMENTALIZED
003 DEPTLIZD (3          ) 010                        8TH GRADE ORGANIZATION:  DEPARTMENTALIZED
004 SCH8ORG? (M          ) 001                        8TH GRADE ORGANIZATION:  MISSING

CONDITIONING VARIABLE ID:  SCHL0005
DESCRIPTION:               ARE 4TH GRADERS ASSIGNED TO CLASSES BY ABILITY?
GRADES/ASSESSMENTS:        N04, S04
CONDITIONING VAR LABEL:    CLA ABL4
NAEP ID:                   C031100                    TOTAL NUMBER OF SPECIFIED CONTRASTS:     3
TYPE OF CONTRAST:          CLASS                      NUMBER OF INDEPENDENT CONTRASTS:         2

001 CLAABL-Y (1          ) 00                         4TH GRADERS ASSIGNED BY ABILITY:  YES
002 CLAABL-N (2          ) 10                         4TH GRADERS ASSIGNED BY ABILITY:  NO
003 CLAABL-? (M          ) 01                         4TH GRADERS ASSIGNED BY ABILITY:  MISSING
```

Table C-2 (continued)

Conditioning Variables Common to the Main Reading, U.S. History, and Geography Assessments

```
CONDITIONING VARIABLE ID:  SCHL0006
DESCRIPTION:               ARE 8TH GRADERS ASSIGNED TO ENGLISH BY ABILITY?
GRADES/ASSESSMENTS:        N08
CONDITIONING VAR LABEL:    ENG ABL8
NAEP ID:                   C034401                    TOTAL NUMBER OF SPECIFIED CONTRASTS:     3
TYPE OF CONTRAST:          CLASS                      NUMBER OF INDEPENDENT CONTRASTS:         2

001 ENGABL-Y (1          ) 00                         ENGLISH BY ABILITY:  YES (8TH GRADE)
002 ENGABL-N (2          ) 10                         ENGLISH BY ABILITY:  NO (8TH GRADE)
003 ENGABL-? (M          ) 01                         ENGLISH BY ABILITY:  MISSING (8TH GRADE)

CONDITIONING VARIABLE ID:  SCHL0007
DESCRIPTION:               ARE 12TH GRADERS ASSIGNED TO ENGLISH BY ABILITY?
GRADES/ASSESSMENTS:        N12
CONDITIONING VAR LABEL:    ENGABL12
NAEP ID:                   C035001                    TOTAL NUMBER OF SPECIFIED CONTRASTS:     3
TYPE OF CONTRAST:          CLASS                      NUMBER OF INDEPENDENT CONTRASTS:         2

001 ENGABL-Y (1          ) 00                         ENGLISH BY ABILITY:  YES (12TH GRADE)
002 ENGABL-N (2          ) 10                         ENGLISH BY ABILITY:  NO (12TH GRADE)
003 ENGABL-? (M          ) 01                         ENGLISH BY ABILITY:  MISSING (12TH GRADE)

CONDITIONING VARIABLE ID:  SCHL0008
DESCRIPTION:               HAS READING BEEN IDENTIFIED AS A PRIORITY? (GRADES 4 AND 8)
GRADES/ASSESSMENTS:        N04, S04, N08
CONDITIONING VAR LABEL:    RD PRIOR
NAEP ID:                   C031601                    TOTAL NUMBER OF SPECIFIED CONTRASTS:     3
TYPE OF CONTRAST:          CLASS                      NUMBER OF INDEPENDENT CONTRASTS:         2

001 RD PRI-Y (1          ) 00                         READING PRIORITY:  YES
002 RD PRI-N (2          ) 10                         READING PRIORITY:  NO
003 RD PRI-? (M          ) 01                         READING PRIORITY:  MISSING

CONDITIONING VARIABLE ID:  SCHL0009
DESCRIPTION:               WHAT PERCENT OF STUDENTS RECEIVE SUBSIDIZED LUNCH?
GRADES/ASSESSMENTS:        N04, S04, N08, N12
CONDITIONING VAR LABEL:    %SUB LUN
NAEP ID:                   C032001                    TOTAL NUMBER OF SPECIFIED CONTRASTS:     6
TYPE OF CONTRAST:          CLASS                      NUMBER OF INDEPENDENT CONTRASTS:         5

001 %SUBLUN1 (1,2,3      ) 00000                      PERCENT SUBSIDIZED LUNCH:  NONE, 1-5%, 6-10%
002 %SUBLUN2 (4          ) 10000                      PERCENT SUBSIDIZED LUNCH:  11-25%
003 %SUBLUN3 (5          ) 01000                      PERCENT SUBSIDIZED LUNCH:  26-50%
004 %SUBLUN4 (6          ) 00100                      PERCENT SUBSIDIZED LUNCH:  51-75%
005 %SUBLUN5 (7,8        ) 00010                      PERCENT SUBSIDIZED LUNCH:  76-90%, OVER 90%
006 %SUBLUN? (M          ) 00001                      PERCENT SUBSIDIZED LUNCH:  MISSING

CONDITIONING VARIABLE ID:  SCHL0010
DESCRIPTION:               WHAT PERCENT OF STUDENTS RECEIVE REMEDIAL READING INSTRUCTION?
GRADES/ASSESSMENTS:        N04, S04, N08, N12
CONDITIONING VAR LABEL:    %REMREAD
NAEP ID:                   C032002                    TOTAL NUMBER OF SPECIFIED CONTRASTS:     6
TYPE OF CONTRAST:          CLASS                      NUMBER OF INDEPENDENT CONTRASTS:         5

001 %REMRED1 (1,2        ) 00000                      PERCENT REMEDIAL READING:  NONE, 1-5%
002 %REMRED2 (3          ) 10000                      PERCENT REMEDIAL READING:  6-10%
003 %REMRED3 (4          ) 01000                      PERCENT REMEDIAL READING:  11-25%
004 %REMRED4 (5,6        ) 00100                      PERCENT REMEDIAL READING:  26-50%, 51-75%
005 %REMRED5 (7,8        ) 00010                      PERCENT REMEDIAL READING:  76-90%, OVER 90%
006 %REMRED? (M          ) 00001                      PERCENT REMEDIAL READING:  MISSING
```

BEST COPY AVAILABLE

Table C-2 (continued)

Conditioning Variables Common to the Main Reading, U.S. History, and Geography Assessments

```
CONDITIONING VARIABLE ID:  SCHL0015
DESCRIPTION:               WHAT PERCENTAGE OF STUDENTS ARE ENROLLED AT BEGINNING AND END OF SCHOOL YEAR?
GRADES/ASSESSMENTS:        N04, S04, N08, N12
CONDITIONING VAR LABEL:    %ENR@EOY
NAEP ID:                   C033700                      TOTAL NUMBER OF SPECIFIED CONTRASTS:     5
TYPE OF CONTRAST:          CLASS                        NUMBER OF INDEPENDENT CONTRASTS:         4

001 %ENREOY1 (1          ) 0000                         PERCENT ENROLLED AT END OR YEAR: 98-100%
002 %ENREOY2 (2          ) 1000                         PERCENT ENROLLED AT END OF YEAR: 95-97%
003 %ENREOY3 (3          ) 0100                         PERCENT ENROLLED AT END OF YEAR: 90-94%
004 %ENREOY4 (4          ) 0010                         PERCENT ENROLLED AT END OF YEAR: LESS THAN 90%
005 %ENREOY? (M          ) 0001                         PERCENT ENROLLED AT END OF YEAR: MISSING

CONDITIONING VARIABLE ID:  SCHL0016
DESCRIPTION:               DOES SCHOOL INVOLVE PARENTS AS AIDES IN CLASS?
GRADES/ASSESSMENTS:        N04, S04, N08, N12
CONDITIONING VAR LABEL:    PAR AIDE
NAEP ID:                   C032207                      TOTAL NUMBER OF SPECIFIED CONTRASTS:     4
TYPE OF CONTRAST:          CLASS                        NUMBER OF INDEPENDENT CONTRASTS:         3

001 PARAID-R (1          ) 000                          PARENTS AS AIDES IN CLASS:  ROUTINELY
002 PARAID-O (2          ) 100                          PARENTS AS AIDES IN CLASS:  OCCASIONALLY
003 PARAID-N (3          ) 010                          PARENTS AS AIDES IN CLASS:  NO
004 PARAID-? (M          ) 001                          PARENTS AS AIDES IN CLASS:  MISSING

CONDITIONING VARIABLE ID:  SCHL0017
DESCRIPTION:               DOES YOUR SCHOOL HAVE PARENTS REVIEW OR SIGN STUDENTS' HOMEWORK?
GRADES/ASSESSMENTS:        N04, S04, N08, N12
CONDITIONING VAR LABEL:    PARREVHW
NAEP ID:                   C032209                      TOTAL NUMBER OF SPECIFIED CONTRASTS:     3
TYPE OF CONTRAST:          CLASS                        NUMBER OF INDEPENDENT CONTRASTS:         2

001 PARHW-RO (1          ) 00                           PARENTS REVIEW HOMEWORK:  YES, ROUTINELY
002 PARHW-OC (2          ) 10                           PARENTS REVIEW HOMEWORK:  YES, OCCASIONALLY
003 PARHW-N? (3,M        ) 01                           PARENTS REVIEW HOMEWORK:  NO, MISSING

CONDITIONING VARIABLE ID:  SCHL0018
DESCRIPTION:               DOES YOUR SCHOOL ASSIGN HOMEWORK FOR STUDENTS TO DO WITH PARENTS?
GRADES/ASSESSMENTS:        N04, S04, N08, N12
CONDITIONING VAR LABEL:    STUPARHW
NAEP ID:                   C032210                      TOTAL NUMBER OF SPECIFIED CONTRASTS:     3
TYPE OF CONTRAST:          CLASS                        NUMBER OF INDEPENDENT CONTRASTS:         2

001 S/PHW-RO (1          ) 00                           STUDENT/PARENT HOMEWORK:  YES, ROUTINELY
002 S/PHW-OC (2          ) 10                           STUDENT/PARENT HOMEWORK:  YES, OCCASIONALLY
003 S/PHW-N? (3,M        ) 01                           STUDENT/PARENT HOMEWORK:  NO, MISSING

CONDITIONING VARIABLE ID:  SCHL0019
DESCRIPTION:               DOES YOUR SCHOOL HAVE A PARENT VOLUNTEER PROGRAM?
GRADES/ASSESSMENTS:        N04, S04, N08, N12
CONDITIONING VAR LABEL:    PARVOLPG
NAEP ID:                   C032211                      TOTAL NUMBER OF SPECIFIED CONTRASTS:     3
TYPE OF CONTRAST:          CLASS                        NUMBER OF INDEPENDENT CONTRASTS:         2

001 VOLPG-RO (1          ) 00                           PARENT VOLUNTEER PROGRAM:  YES, ROUTINELY
002 VOLPG-OC (2          ) 10                           PARENT VOLUNTEER PROGRAM:  YES, OCCASIONALLY
003 VOLPG-N? (3,M        ) 01                           PARENT VOLUNTEER PROGRAM:  NO, MISSING

CONDITIONING VARIABLE ID:  SCHL0020
DESCRIPTION:               DOES YOUR SCHOOL RECEIVE CHAPTER 1 FUNDING?
GRADES/ASSESSMENTS:        N04, S04, N08, N12
CONDITIONING VAR LABEL:    SCHCHAP1
NAEP ID:                   C036701                      TOTAL NUMBER OF SPECIFIED CONTRASTS:     3
TYPE OF CONTRAST:          CLASS                        NUMBER OF INDEPENDENT CONTRASTS:         2
```

Table C-2 (continued)

Conditioning Variables Common to the Main Reading, U.S. History, and Geography Assessments

```
001 SCHCH1-Y (1          ) 00                SCHOOL CHAPTER 1 FUNDING:  YES
002 SCHCH1-N (2          ) 10                SCHOOL CHAPTER 1 FUNDING:  NO
003 SCHCH1-? (M          ) 01                SCHOOL CHAPTER 1 FUNDING:  MISSING

CONDITIONING VARIABLE ID:  SCHL0021
DESCRIPTION:               WHAT PERCENTAGE OF STUDENTS IN YOUR SCHOOL ARE CHAPTER 1 ELIGIBLE?
GRADES/ASSESSMENTS:        N04, S04, N08, N12
CONDITIONING VAR LABEL:    SCH%CHP1
NAEP ID:                   C036801           TOTAL NUMBER OF SPECIFIED CONTRASTS:    6
TYPE OF CONTRAST:          CLASS             NUMBER OF INDEPENDENT CONTRASTS:        5

001 %CHAP1-1 (1          ) 00000             PERCENT CHAPTER 1 ELIGIBLE STUDENTS:  10% OR BELOW
002 %CHAP1-2 (2          ) 10000             PERCENT CHAPTER 1 ELIGIBLE STUDENTS:  11-25%
003 %CHAP1-3 (3          ) 01000             PERCENT CHAPTER 1 ELIGIBLE STUDENTS:  26-75%
004 %CHAP1-4 (4          ) 00100             PERCENT CHAPTER 1 ELIGIBLE STUDENTS:  76-99%
005 %CHAP1-5 (5          ) 00010             PERCENT CHAPTER 1 ELIGIBLE STUDENTS:  100%
006 %CHAP1-? (M          ) 00001             PERCENT CHAPTER 1 ELIGIBLE STUDENTS:  MISSING

CONDITIONING VARIABLE ID:  SCHL0022
DESCRIPTION:               PERCENT OF STUDENTS FROM A RURAL AREA (MISSING VS NON-MISSING)
GRADES/ASSESSMENTS:        N04, S04, N08, N12
CONDITIONING VAR LABEL:    SCH%RURL
NAEP ID:                   C036201           TOTAL NUMBER OF SPECIFIED CONTRASTS:    2
TYPE OF CONTRAST:          CLASS             NUMBER OF INDEPENDENT CONTRASTS:        1

001 SCH%RURL (M          ) 0                 PERCENT OF STUDENTS - RURAL:  MISSING
002 SCH%RURL (@          ) 1                 PERCENT OF STUDENTS - RURAL:  NOT MISSING

CONDITIONING VARIABLE ID:  SCHL0023
DESCRIPTION:               PERCENT OF STUDENTS FROM A RURAL AREA OF LESS THAN 2,500 (LINEAR)
GRADES/ASSESSMENTS:        N04, S04, N08, N12
CONDITIONING VAR LABEL:    RUR%-LIN
NAEP ID:                   C036201           TOTAL NUMBER OF SPECIFIED CONTRASTS:    1
TYPE OF CONTRAST:          LINEAR            NUMBER OF INDEPENDENT CONTRASTS:        1

001 RUR%-LIN (0-100,M=0  )  0.0 +  1.0*X     PERCENT OF STUDENTS FROM A RURAL AREA OF LESS THAN
                                             2,500 (LINEAR)

CONDITIONING VARIABLE ID:  SCHL0026
DESCRIPTION:               PERCENT OF STUDENTS FROM A CITY (MISSING VS NON-MISSING)
GRADES/ASSESSMENTS:        N04, S04, N08, N12
CONDITIONING VAR LABEL:    SCH%CITY
NAEP ID:                   C036203           TOTAL NUMBER OF SPECIFIED CONTRASTS:    2
TYPE OF CONTRAST:          CLASS             NUMBER OF INDEPENDENT CONTRASTS:        1

001 SCH%CITY (M          ) 0                 PERCENT OF STUDENTS - CITY:  MISSING
002 SCH%CITY (@          ) 1                 PERCENT OF STUDENTS - CITY:  NOT MISSING

CONDITIONING VARIABLE ID:  SCHL0027
DESCRIPTION:               PERCENT OF STUDENTS FROM A TOWN OF 10,000 OR MORE (LINEAR)
GRADES/ASSESSMENTS:        N04, S04, N08, N12
CONDITIONING VAR LABEL:    CITY%-LN
NAEP ID:                   C036203           TOTAL NUMBER OF SPECIFIED CONTRASTS:    1
TYPE OF CONTRAST:          LINEAR            NUMBER OF INDEPENDENT CONTRASTS:        1

001 CITY%-LN (0-100,M=0  )  0.0 +  1.0*X     PERCENT OF STUDENTS FROM A TOWN OF 10,000 OR MORE
                                             (LINEAR)

CONDITIONING VARIABLE ID:  SCHL0028
DESCRIPTION:               PERCENT OF STUDENTS WITH PROFESSIONAL PARENTS (MISSING VS NON-MISSING)
GRADES/ASSESSMENTS:        N04, S04, N08, N12
CONDITIONING VAR LABEL:    PAR%PROF
NAEP ID:                   C036301           TOTAL NUMBER OF SPECIFIED CONTRASTS:    2
TYPE OF CONTRAST:          CLASS             NUMBER OF INDEPENDENT CONTRASTS:        1
```

Table C-2 (continued)

Conditioning Variables Common to the Main Reading, U.S. History, and Geography Assessments

```
001 PAR%PROF (M          ) 0                   PERCENT OF STUDENTS - PARENTS IN PROFESSIONS:
                                                                     MISSING
002 PAR%PROF (@          ) 1                   PERCENT OF STUDENTS - PARENTS IN PROFESSIONS:  NOT
                                                                     MISSING

CONDITIONING VARIABLE ID:  SCHL0029
DESCRIPTION:               PERCENT OF STUDENTS WITH PARENTS IN PROFESSIONAL/MANAGEMENT POSITIONS (LINEAR)
GRADES/ASSESSMENTS:        N04, S04, N08, N12
CONDITIONING VAR LABEL:    PROF%-LN
NAEP ID:                   C036301             TOTAL NUMBER OF SPECIFIED CONTRASTS:     1
TYPE OF CONTRAST:          LINEAR              NUMBER OF INDEPENDENT CONTRASTS:         1

001 PROF%-LN (0-100,M=0  )  0.0 +  1.0*X       PERCENT OF STUDENTS WITH PARENTS IN
                                          PROFESSIONAL/MANAGEMENT POSITIONS (LINEAR)

CONDITIONING VARIABLE ID:  SCHL0032
DESCRIPTION:               PERCENT OF STUDENTS WITH PARENTS IN BLUE-COLLAR POSITIONS (MISSING VS NON-MISSING)
GRADES/ASSESSMENTS:        N04, S04, N08, N12
CONDITIONING VAR LABEL:    PAR%BCOL
NAEP ID:                   C036303             TOTAL NUMBER OF SPECIFIED CONTRASTS:     2
TYPE OF CONTRAST:          CLASS               NUMBER OF INDEPENDENT CONTRASTS:         1

001 PAR%BCOL (M          ) 0                   PERCENT OF STUDENTS - PARENTS IN BLUE-COLLAR
                                                                     POSITIONS:  MISSING
002 PAR%BCOL (@          ) 1                   PERCENT OF STUDENTS - PARENTS IN BLUE-COLLAR
                                                                     POSITIONS:  NOT MISSING

CONDITIONING VARIABLE ID:  SCHL0033
DESCRIPTION:               PERCENT OF STUDENTS WITH PARENTS IN BLUE-COLLAR POSITIONS (LINEAR)
GRADES/ASSESSMENTS:        N04, S04, N08, N12
CONDITIONING VAR LABEL:    BCOL%-LN
NAEP ID:                   C036303             TOTAL NUMBER OF SPECIFIED CONTRASTS:     1
TYPE OF CONTRAST:          LINEAR              NUMBER OF INDEPENDENT CONTRASTS:         1

001 BCOL%-LN (0-100 M=0  )  0.0 +  1.0*X       PERCENT OF STUDENTS WITH PARENTS IN BLUE-COLLAR
                                               POSITIONS (LINEAR)

CONDITIONING VARIABLE ID:  SCHL0034
DESCRIPTION:               PERCENT OF STUDENTS WITH PARENTS WHO ARE FARM WORKERS (MISSING VS NON-MISSING)
GRADES/ASSESSMENTS:        N04, S04, N08, N12
CONDITIONING VAR LABEL:    PAR%FARM
NAEP ID:                   C036304             TOTAL NUMBER OF SPECIFIED CONTRASTS:     2
TYPE OF CONTRAST:          CLASS               NUMBER OF INDEPENDENT CONTRASTS:         1

001 PAR%FARM (M          ) 0                   PERCENT OF STUDENTS - PARENTS WHO ARE FARM
                                                                     WORKERS:  MISSING
002 PAR%FARM (@          ) 1                   PERCENT OF STUDENTS - PARENTS WHO ARE FARM
                                                                     WORKERS:  NOT MISSING

CONDITIONING VARIABLE ID:  SCHL0035
DESCRIPTION:               PERCENT OF STUDENTS WITH PARENTS WHO ARE FARM WORKERS (LINEAR)
GRADES/ASSESSMENTS:        N04, S04, N08, N12
CONDITIONING VAR LABEL:    FARM%-LN
NAEP ID:                   C036304             TOTAL NUMBER OF SPECIFIED CONTRASTS:     1
TYPE OF CONTRAST:          LINEAR              NUMBER OF INDEPENDENT CONTRASTS:         1

001 FARM%-LN (0-100,M=0  )  0.0 +  1.0*X       PERCENT OF STUDENTS WITH PARENTS WHO ARE FARM
                                               WORKERS (LINEAR)

CONDITIONING VARIABLE ID:  SCHL0036
DESCRIPTION:               PERCENT OF STUDENTS WITH PARENTS WHO ARE IRREGULARLY EMPLOYED (MISSING VS
                           NON-MISSING)
GRADES/ASSESSMENTS:        N04, S04, N08, N12
CONDITIONING VAR LABEL:    PAR%IRRE
NAEP ID:                   C036305             TOTAL NUMBER OF SPECIFIED CONTRASTS:     2
```

Table C-2 (continued)

Conditioning Variables Common to the Main Reading, U.S. History, and Geography Assessments

```
TYPE OF CONTRAST:          CLASS                          NUMBER OF INDEPENDENT CONTRASTS:        1

001 PAR%IRRE (M         ) 0                               PERCENT OF STUDENTS - PARENTS WHO ARE IRREGULARLY
                                                                                EMPLOYED:  MISSING
002 PAR%IRRE (@         ) 1                               PERCENT OF STUDENTS - PARENTS WHO ARE IRREGULARLY
                                                                                EMPLOYED:  NOT MISSING

CONDITIONING VARIABLE ID:  SCHL0037
DESCRIPTION:               PERCENT OF STUDENTS WITH PARENTS WHO ARE IRREGULARLY EMPLOYED BUT NOT ON WELFARE
                           (LINEAR)
GRADES/ASSESSMENTS:        N04, S04, N08, N12
CONDITIONING VAR LABEL:    IRRE%-LN
NAEP ID:                   C036305                        TOTAL NUMBER OF SPECIFIED CONTRASTS:    1
TYPE OF CONTRAST:          LINEAR                         NUMBER OF INDEPENDENT CONTRASTS:        1

001 IRRE%-LN (0-100,M=0  )  0.0 +  1.0*X                  PERCENT OF STUDENTS WITH PARENTS WHO ARE
                                                          IRREGULARLY EMPLOYED BUT NOT ON WELFARE (LINEAR)

CONDITIONING VARIABLE ID:  SCHL0038
DESCRIPTION:               PERCENT OF STUDENTS WITH PARENTS WHO ARE WELFARE RECIPIENTS (MISSING VS
                           NON-MISSING)
GRADES/ASSESSMENTS:        N04, S04, N08, N12
CONDITIONING VAR LABEL:    PAR%WELF
NAEP ID:                   C036306                        TOTAL NUMBER OF SPECIFIED CONTRASTS:    2
TYPE OF CONTRAST:          CLASS                          NUMBER OF INDEPENDENT CONTRASTS:        1

001 PAR%WELF (M         ) 0                               PERCENT OF STUDENTS - PARENTS WHO ARE WELFARE
                                                                                RECIPIENTS:  MISSING
002 PAR%WELF (@         ) 1                               PERCENT OF STUDENTS - PARENTS WHO ARE WELFARE
                                                                                RECIPIENTS:  NOT MISSING

CONDITIONING VARIABLE ID:  SCHL0039
DESCRIPTION:               PERCENT OF STUDENTS WITH PARENTS WHO ARE WELFARE RECIPIENTS (LINEAR)
GRADES/ASSESSMENTS:        N04, S04, N08, N12
CONDITIONING VAR LABEL:    WELF%-LN
NAEP ID:                   C036306                        TOTAL NUMBER OF SPECIFIED CONTRASTS:    1
TYPE OF CONTRAST:          LINEAR                         NUMBER OF INDEPENDENT CONTRASTS:        1

001 WELF%-LN (0-100,M=0  )  0.0 +  1.0*X                  PERCENT OF STUDENTS WITH PARENTS WHO ARE WELFARE
                                                          RECIPIENTS (LINEAR)

CONDITIONING VARIABLE ID:  SCHL0042
DESCRIPTION:               WHAT IS THE PRIMARY WAY YOUR LIBRARY IS STAFFED?
GRADES/ASSESSMENTS:        N04, S04, N08, N12
CONDITIONING VAR LABEL:    LIBSTAFF
NAEP ID:                   C036601                        TOTAL NUMBER OF SPECIFIED CONTRASTS:    5
TYPE OF CONTRAST:          CLASS                          NUMBER OF INDEPENDENT CONTRASTS:        4

001 NO LIBRY (1         ) 0000                            LIBRARY STAFFING:  NO LIBRARY IN SCHOOL
002 LSTAF-NO (2         ) 1000                            LIBRARY STAFFING:  LIBRARY IN SCHOOL BUT
                                                                             NO/VOLUNTEER STAFF
003 LSTAF-PT (3         ) 0100                            LIBRARY STAFFING:  PART-TIME STAFF
004 LSTAF-FT (4         ) 0010                            LIBRARY STAFFING:  FULL-TIME STAFF
005 LSTAF-?  (M         ) 0001                            LIBRARY STAFFING:  MISSING

CONDITIONING VARIABLE ID:  SCHL0043
DESCRIPTION:               WHO TEACHES ENGLISH/LANGUAGE ARTS TO 8TH GRADE?
GRADES/ASSESSMENTS:        N08
CONDITIONING VAR LABEL:    #TCHSUBJ
NAEP ID:                   C034701                        TOTAL NUMBER OF SPECIFIED CONTRASTS:    4
TYPE OF CONTRAST:          CLASS                          NUMBER OF INDEPENDENT CONTRASTS:        3
```

Table C-2 (continued)

Conditioning Variables Common to the Main Reading, U.S. History, and Geography Assessments

```
001 TCHSUB>1 (1          ) 000                 8TH-GRADE ENGLISH:  TEACHERS WITH MORE THAN ONE
                                                                   SUBJECT
002 TCHSUB=1 (2          ) 100                 8TH-GRADE ENGLISH:  TEACHERS WITH ONE SUBJECT
003 NO ENGL8 (3          ) 010                 8TH-GRADE ENGLISH:  SUBJECT NOT TAUGHT
004 TCHSUB-? (M          ) 001                 8TH-GRADE ENGLISH:  MISSING

CONDITIONING VARIABLE ID:  SCHL0044
DESCRIPTION:               NUMBER OF STUDENTS ENROLLED IN AP ENGLISH
GRADES/ASSESSMENTS:        N12
CONDITIONING VAR LABEL:    #AP ENGL
NAEP ID:                   C035802             TOTAL NUMBER OF SPECIFIED CONTRASTS:     7
TYPE OF CONTRAST:          CLASS               NUMBER OF INDEPENDENT CONTRASTS:         6

001 #AP ENG1 (1          ) 000000              # STUDENTS AP ENGLISH:  NONE
002 #AP ENG2 (2          ) 100000              # STUDENTS AP ENGLISH:  1-5
003 #AP ENG3 (3          ) 010000              # STUDENTS AP ENGLISH:  6-10
004 #AP ENG4 (4          ) 001000              # STUDENTS AP ENGLISH:  11-25
005 #AP ENG5 (5          ) 000100              # STUDENTS AP ENGLISH:  26-50
006 #AP ENG6 (6          ) 000010              # STUDENTS AP ENGLISH:  MORE THAN 50
007 #AP ENG? (M          ) 000001              # STUDENTS AP ENGLISH:  MISSING

CONDITIONING VARIABLE ID:  SCHL0045
DESCRIPTION:               NUMBER OF SEMESTERS ENGLISH/LITERATURE/WRITING REQUIRED (GRADE 12)
GRADES/ASSESSMENTS:        N12
CONDITIONING VAR LABEL:    #REQ ENG
NAEP ID:                   C035201             TOTAL NUMBER OF SPECIFIED CONTRASTS:     3
TYPE OF CONTRAST:          CLASS               NUMBER OF INDEPENDENT CONTRASTS:         2

001 #REQENG- (1-8        ) 00                  SEMESTERS ENGLISH:  NONE, ONE, TWO, ..., SEVEN
002 #REQENG8 (9          ) 10                  SEMESTERS ENGLISH:  EIGHT
003 #REQENG? (M          ) 01                  SEMESTERS ENGLISH:  MISSING

CONDITIONING VARIABLE ID:  SCHL0047
DESCRIPTION:               ARE COMPUTERS AVAILIABE AT ALL TIMES IN CLASSROOMS?
GRADES/ASSESSMENTS:        N04, S04, N08, N12
CONDITIONING VAR LABEL:    COMP CLS
NAEP ID:                   C035701             TOTAL NUMBER OF SPECIFIED CONTRASTS:     2
TYPE OF CONTRAST:          CLASS               NUMBER OF INDEPENDENT CONTRASTS:         1

001 CMPCLS-Y (1          ) 0                   COMPUTERS AVAILABLE IN CLASS:  YES
002 CMPCL-N? (2,M        ) 1                   COMPUTERS AVIALABLE IN CLASS:  NO, MISSING

CONDITIONING VARIABLE ID:  SCHL0048
DESCRIPTION:               ARE COMPUTERS GROUPED IN A SEPARATE COMPUTER LABORATORY AVAILABLE TO CLASSES?
GRADES/ASSESSMENTS:        N04, S04, N08, N12
CONDITIONING VAR LABEL:    COMP LAB
NAEP ID:                   C035702             TOTAL NUMBER OF SPECIFIED CONTRASTS:     2
TYPE OF CONTRAST:          CLASS               NUMBER OF INDEPENDENT CONTRASTS:         1

001 CMPLAB-Y (1          ) 0                   COMPUTERS IN A LAB:  YES
002 CMPLB-N? (2,M        ) 1                   COMPUTERS IN A LAB:  NO, MISSING

CONDITIONING VARIABLE ID:  SCHL0049
DESCRIPTION:               ARE COMPUTERS AVAILABLE TO BRING TO CLASSROOMS WHEN NEEDED?
GRADES/ASSESSMENTS:        N04, S04, N08, N12
CONDITIONING VAR LABEL:    COMP BRG
NAEP ID:                   C035703             TOTAL NUMBER OF SPECIFIED CONTRASTS:     2
TYPE OF CONTRAST:          CLASS               NUMBER OF INDEPENDENT CONTRASTS:         1

001 CMPBRG-Y (1          ) 0                   BRING COMPUTERS TO CLASS:  YES
002 CMPBR-N? (2,M        ) 1                   BRING COMPUTERS TO CLASS:  NO, MISSING
```

Table C-3

Conditioning Variables Common to the U.S. History and Geography Assessments

```
CONDITIONING VARIABLE ID:  SCHL0050
DESCRIPTION:               HAS HISTORY BEEN IDENTIFIED AS A PRIORITY? (GRADES 4 AND 8)
GRADES/ASSESSMENTS:        N04, N08
CONDITIONING VAR LABEL:    HS PRIOR
NAEP ID:                   C031608                TOTAL NUMBER OF SPECIFIED CONTRASTS:    3

TYPE OF CONTRAST:          CLASS                  NUMBER OF INDEPENDENT CONTRASTS:        2


001 HS PRI-Y (1          ) 00HISTORY PRIORITY:  YES
002 HS PRI-N (2          ) 10HISTORY PRIORITY:  NO
003 HS PRI-? (M          ) 01HISTORY PRIORITY:  MISSING

CONDITIONING VARIABLE ID:  SCHL0051
DESCRIPTION:               HAS GEOGRAPHY BEEN IDENTIFIED AS A PRIORITY? (GRADES 4 AND 8)
GRADES/ASSESSMENTS:        N04, N08
CONDITIONING VAR LABEL:    GE PRIOR
NAEP ID:                   C031609                TOTAL NUMBER OF SPECIFIED CONTRASTS:    3

TYPE OF CONTRAST:          CLASS                  NUMBER OF INDEPENDENT CONTRASTS:        2

001 GE PRI-Y (1          ) 00GEOGRAPHY PRIORITY: YES
002 GE PRI-N (2          ) 10GEOGRAPHY PRIORITY: NO
003 GE PRI-? (M          ) 01GEOGRAPHY PRIORITY: MISSING

CONDITIONING VARIABLE ID:  SCHL0052
DESCRIPTION:               WHO TEACHES U. S. HISTORY TO 8TH GRADE?
GRADES/ASSESSMENTS:        N08
CONDITIONING VAR LABEL:    HTSUBJ#
NAEP ID:                   C034704                TOTAL NUMBER OF SPECIFIED CONTRASTS:    4
TYPE OF CONTRAST:          CLASS                  NUMBER OF INDEPENDENT CONTRASTS:        3


001 HTSUB#>1 (1          ) 000                    8TH-GRADE U. S. HISTORY: TEACHERS WITH MORE THAN
                                                                           ONE SUBJECT
002 HTSUB#=1 (2          ) 100                    8TH-GRADE U. S. HISTORY: TEACHERS WITH ONE SUBJECT

003 NO HIST8 (3          ) 010                    8TH-GRADE U. S. HISTORY: SUBJECT NOT TAUGHT

004 HTSUB#-? (M          ) 001                    8TH-GRADE U. S. HISTORY: MISSING

CONDITIONING VARIABLE ID:  SCHL0053
DESCRIPTION:               WHO TEACHES GEOGRAPHY TO 8TH GRADE?
GRADES/ASSESSMENTS:        N08
CONDITIONING VAR LABEL:    GTSUBJ#
NAEP ID:                   C034705                TOTAL NUMBER OF SPECIFIED CONTRASTS:    4

TYPE OF CONTRAST:          CLASS                  NUMBER OF INDEPENDENT CONTRASTS:        3


001 GTSUB#>1 (1          ) 000                    8TH-GRADE GEOGRAPHY: TEACHERS WITH MORE THAN ONE
                                                                       SUBJECT

002 GTSUB#=1 (2          ) 100                    8TH-GRADE GEOGRAPHY: TEACHERS WITH ONE SUBJECT

003 NO GEOG8 (3          ) 010                    8TH-GRADE GEOGRAPHY: SUBJECT NOT TAUGHT

004 GTSUB#-? (M          ) 001                    8TH-GRADE GEOGRAPHY: MISSING

CONDITIONING VARIABLE ID:  SCHL0054
DESCRIPTION:               ARE 8TH GRADERS ASSIGNED TO HISTORY BY ABILITY?
GRADES/ASSESSMENTS:        N08
```

BEST COPY AVAILABLE

Table C-3 (continued)

Conditioning Variables Common to the U.S. History and Geography Assessments

```
CONDITIONING VAR LABEL:     HIS ABL8
NAEP ID:                    C034404                      TOTAL NUMBER OF SPECIFIED CONTRASTS:     3
TYPE OF CONTRAST:           CLASS                        NUMBER OF INDEPENDENT CONTRASTS:         2

001 HISABL-Y (1          ) 00HISTORY BY ABILITY:  YES (8TH GRADE)
002 HISABL-N (2          ) 10HISTORY BY ABILITY:  NO (8TH GRADE)
003 HISABL-? (M          ) 01HISTORY BY ABILITY:  MISSING (8TH GRADE)

CONDITIONING VARIABLE ID:   SCHL0055
DESCRIPTION:                ARE 8TH GRADERS ASSIGNED TO GEOGRAPHY BY ABILITY?
GRADES/ASSESSMENTS:         N08
CONDITIONING VAR LABEL:     GEO ABL8
NAEP ID:                    C034405                      TOTAL NUMBER OF SPECIFIED CONTRASTS:     3

TYPE OF CONTRAST:           CLASS                        NUMBER OF INDEPENDENT CONTRASTS:         2

001 GEOABL-Y (1          ) 00GEOGRAPHY BY ABILITY: YES (8TH GRADE)
002 GEOABL-N (2          ) 10GEOGRAPHY BY ABILITY: NO (8TH GRADE)
003 GEOABL-? (M          ) 01GEOGRAPHY BY ABILITY: MISSING (8TH GRADE)

CONDITIONING VARIABLE ID:   SCHL0056
DESCRIPTION:                ARE 12TH GRADERS ASSIGNED TO HISTORY BY ABILITY?
GRADES/ASSESSMENTS:         N12
CONDITIONING VAR LABEL:     HISABL12
NAEP ID:                    C035004                      TOTAL NUMBER OF SPECIFIED CONTRASTS:     3

TYPE OF CONTRAST:           CLASS                        NUMBER OF INDEPENDENT CONTRASTS:         2

001 HISABL-Y (1          ) 00HISTORY BY ABILITY:  YES (12TH GRADE)
002 HISABL-N (2          ) 10HISTORY BY ABILITY:  NO (12TH GRADE)
003 HISABL-? (M          ) 01HISTORY BY ABILITY:  MISSING (12TH GRADE)

CONDITIONING VARIABLE ID:   SCHL0057
DESCRIPTION:                ARE 12TH GRADERS ASSIGNED TO GEOGRAPHY BY ABILITY?
GRADES/ASSESSMENTS:         N12
CONDITIONING VAR LABEL:     GEOABL12
NAEP ID:                    C035005                      TOTAL NUMBER OF SPECIFIED CONTRASTS:     3

TYPE OF CONTRAST:           CLASS                        NUMBER OF INDEPENDENT CONTRASTS:         2

001 GEOABL-Y (1          ) 00GEOGRAPHY BY ABILITY: YES (12TH GRADE)
002 GEOABL-N (2          ) 10GEOGRAPHY BY ABILITY: NO (12TH GRADE)
003 GEOABL-? (M          ) 01GEOGRAPHY BY ABILITY: MISSING (12TH GRADE)

CONDITIONING VARIABLE ID:   SCHL0058
DESCRIPTION:                NUMBER OF SEMESTERS SOCIAL STUDIES REQUIRED (GRADE 12)
GRADES/ASSESSMENTS:         N12
CONDITIONING VAR LABEL:     #REQ SS
NAEP ID:                    C035207                      TOTAL NUMBER OF SPECIFIED CONTRASTS:    10

TYPE OF CONTRAST:           CLASS                        NUMBER OF INDEPENDENT CONTRASTS:         9

001 #REQ_SS0 (1          ) 000000000                     SEMESTERS SOCIAL STUDIES: NONE
002 #REQ_SS1 (2          ) 100000000                     SEMESTERS SOCIAL STUDIES: ONE
003 #REQ_SS2 (3          ) 010000000                     SEMESTERS SOCIAL STUDIES: TWO
004 #REQ_SS3 (4          ) 001000000                     SEMESTERS SOCIAL STUDIES: THREE
005 #REQ_SS4 (5          ) 000100000                     SEMESTERS SOCIAL STUDIES: FOUR
006 #REQ_SS5 (6          ) 000010000                     SEMESTERS SOCIAL STUDIES: FIVE
007 #REQ_SS6 (7          ) 000001000                     SEMESTERS SOCIAL STUDIES: SIX
008 #REQ_SS7 (8          ) 000000100                     SEMESTERS SOCIAL STUDIES: SEVEN
009 #REQ_SS8 (9          ) 000000010                     SEMESTERS SOCIAL STUDIES: EIGHT
010 #REQ_SS? (M          ) 000000001                     SEMESTERS SOCIAL STUDIES: MISSING
```

Table C-3 (continued)

Conditioning Variables Common to the U.S. History and Geography Assessments

```
CONDITIONING VARIABLE ID:  SCHL0059
DESCRIPTION:               NUMBER OF SEMESTERS HISTORY REQUIRED (GRADE 12)
GRADES/ASSESSMENTS:        N12
CONDITIONING VAR LABEL:    #REQ HIS
NAEP ID:                   C035208                        TOTAL NUMBER OF SPECIFIED CONTRASTS:    10
TYPE OF CONTRAST:          CLASS                          NUMBER OF INDEPENDENT CONTRASTS:         9

001 #REQ_HS0 (1          ) 000000000                      SEMESTERS HISTORY: NONE
002 #REQ_HS1 (2          ) 100000000                      SEMESTERS HISTORY: ONE
003 #REQ_HS2 (3          ) 010000000                      SEMESTERS HISTORY: TWO
004 #REQ_HS3 (4          ) 001000000                      SEMESTERS HISTORY: THREE
005 #REQ_HS4 (5          ) 000100000                      SEMESTERS HISTORY: FOUR
006 #REQ_HS5 (6          ) 000010000                      SEMESTERS HISTORY: FIVE
007 #REQ_HS6 (7          ) 000001000                      SEMESTERS HISTORY: SIX
008 #REQ_HS7 (8          ) 000000100                      SEMESTERS HISTORY: SEVEN
009 #REQ_HS8 (9          ) 000000010                      SEMESTERS HISTORY: EIGHT
010 #REQ_HS? (M          ) 000000001                      SEMESTERS HISTORY: MISSING

CONDITIONING VARIABLE ID:  SCHL0060
DESCRIPTION:               NUMBER OF SEMESTERS GEOGRAPHY REQUIRED (GRADE 12)
GRADES/ASSESSMENTS:        N12
CONDITIONING VAR LABEL:    #REQ GEO
NAEP ID:                   C035209                        TOTAL NUMBER OF SPECIFIED CONTRASTS:    10
TYPE OF CONTRAST:          CLASS                          NUMBER OF INDEPENDENT CONTRASTS:         9

001 #REQ_GE0 (1          ) 000000000                      SEMESTERS GEOGRAPHY: NONE
002 #REQ_GE1 (2          ) 100000000                      SEMESTERS GEOGRAPHY: ONE
003 #REQ_GE2 (3          ) 010000000                      SEMESTERS GEOGRAPHY: TWO
004 #REQ_GE3 (4          ) 001000000                      SEMESTERS GEOGRAPHY: THREE
005 #REQ_GE4 (5          ) 000100000                      SEMESTERS GEOGRAPHY: FOUR
006 #REQ_GE5 (6          ) 000010000                      SEMESTERS GEOGRAPHY: FIVE
007 #REQ_GE6 (7          ) 000001000                      SEMESTERS GEOGRAPHY: SIX
008 #REQ_GE7 (8          ) 000000100                      SEMESTERS GEOGRAPHY: SEVEN
009 #REQ_GE8 (9          ) 000000010                      SEMESTERS GEOGRAPHY: EIGHT
010 #REQ_GE? (M          ) 000000001                      SEMESTERS GEOGRAPHY: MISSING

CONDITIONING VARIABLE ID:  SCHL0061
DESCRIPTION:               ARE GEOGRAPHY COURSES OF AT LEAST ONE SEMESTER IN LENGTH TAUGHT IN YOUR SCHOOL?
GRADES/ASSESSMENTS:        N12
CONDITIONING VAR LABEL:    >1SMGEO
NAEP ID:                   C037003                        TOTAL NUMBER OF SPECIFIED CONTRASTS:     3
TYPE OF CONTRAST:          CLASS                          NUMBER OF INDEPENDENT CONTRASTS:         2

001 >1SMGE-Y (1          ) 00AT LEAST ONE SEMESTER GEOGRAPHY COURSES OFFERED:  YES
002 >1SMGE-N (2          ) 10AT LEAST ONE SEMESTER GEOGRAPHY COURSES OFFERED:  NO
003 >1SMGE-? (M          ) 01AT LEAST ONE SEMESTER GEOGRAPHY COURSES OFFERED:  MISSING

CONDITIONING VARIABLE ID:  SCHL0062
DESCRIPTION:               ARE U.S. HISTORY COURSES OF AT LEAST ONE SEMESTER IN LENGTH TAUGHT IN YOUR SCHOOL?
GRADES/ASSESSMENTS:        N12
CONDITIONING VAR LABEL:    >1SMHIS
NAEP ID:                   C037011                        TOTAL NUMBER OF SPECIFIED CONTRASTS:     3
TYPE OF CONTRAST:          CLASS                          NUMBER OF INDEPENDENT CONTRASTS:         2

001 >1SMHS-Y (1          ) 00AT LEAST ONE SEMESTER U.S. HISTORY COURSES OFFERED: YES
002 >1SMHS-N (2          ) 10AT LEAST ONE SEMESTER U.S. HISTORY COURSES OFFERED: NO
003 >1SMHS-? (M          ) 01AT LEAST ONE SEMESTER U.S. HISTORY COURSES OFFERED: MISSING

CONDITIONING VARIABLE ID:  SCHL0063
DESCRIPTION:               NUMBER OF STUDENTS ENROLLED IN AP HISTORY
GRADES/ASSESSMENTS:        N12
CONDITIONING VAR LABEL:    #AP HIST
NAEP ID:                   C035804                        TOTAL NUMBER OF SPECIFIED CONTRASTS:     7
```

Table C-3 (continued)

Conditioning Variables Common to the U.S. History and Geography Assessments

```
TYPE OF CONTRAST:          CLASS                         NUMBER OF INDEPENDENT CONTRASTS:        6

001 #AP HIS1 (1          ) 000000                        # STUDENTS AP HISTORY:  NONE
002 #AP HIS2 (2          ) 100000                        # STUDENTS AP HISTORY:  1-5
003 #AP HIS3 (3          ) 010000                        # STUDENTS AP HISTORY:  6-10
004 #AP HIS4 (4          ) 001000                        # STUDENTS AP HISTORY:  11-25
005 #AP HIS5 (5          ) 000100                        # STUDENTS AP HISTORY:  26-50
006 #AP HIS6 (6          ) 000010                        # STUDENTS AP HISTORY:  MORE THAN 50
007 #AP HIS7 (M          ) 000001                        # STUDENTS AP HISTORY:  MISSING
```

Table C-4

Conditioning Variables Specific to the Main Reading Assessment

```
CONDITIONING VARIABLE ID:  SUBJ0001
DESCRIPTION:               DURING THE PAST MONTH, HOW MANY BOOKS HAVE YOU READ ON YOUR OWN OUTSIDE OF SCHOOL?
GRADES/ASSESSMENTS:        N04, S04, N08, N12
CONDITIONING VAR LABEL:    NBOOKSRD
NAEP ID:                   R810801                          TOTAL NUMBER OF SPECIFIED CONTRASTS:     5
TYPE OF CONTRAST:          CLASS                            NUMBER OF INDEPENDENT CONTRASTS:         4

001 NBOOK-0  (1          ) 0000                             NUMBER OF BOOKS READ:  NONE
002 NBOOK-12 (2          ) 1000                             NUMBER OF BOOKS READ:  ONE OR TWO
003 NBOOK-34 (3          ) 0100                             NUMBER OF BOOKS READ:  THREE OR FOUR
004 NBOOK-5+ (4          ) 0010                             NUMBER OF BOOKS READ:  FIVE OR MORE
005 NBOOK-?  (M          ) 0001                             NUMBER OF BOOKS READ:  MISSING

CONDITIONING VARIABLE ID:  SUBJ0002
DESCRIPTION:               WHAT KIND OF READER DO YOU THINK YOU ARE?
GRADES/ASSESSMENTS:        N04, S04, N08, N12
CONDITIONING VAR LABEL:    KIND RDR
NAEP ID:                   R810201                          TOTAL NUMBER OF SPECIFIED CONTRASTS:     5
TYPE OF CONTRAST:          CLASS                            NUMBER OF INDEPENDENT CONTRASTS:         4

001 VGD RDR  (1          ) 0000                             A VERY GOOD READER
002 GOOD RDR (2          ) 1000                             A GOOD READER
003 AVG RDR  (3          ) 0100                             AN AVERAGE READER
004 POOR RDR (4          ) 0010                             A POOR READER
005 RDR-MISS (M          ) 0001                             KIND OF READER:  MISSING

CONDITIONING VARIABLE ID:  SUBJ0003
DESCRIPTION:               HOW OFTEN DO YOU READ FOR FUN ON YOUR OWN TIME?
GRADES/ASSESSMENTS:        N04, S04, N08, N12
CONDITIONING VAR LABEL:    READ4FUN
NAEP ID:                   R810901                          TOTAL NUMBER OF SPECIFIED CONTRASTS:     5
TYPE OF CONTRAST:          CLASS                            NUMBER OF INDEPENDENT CONTRASTS:         4

001 RD4FUN-1 (1          ) 0000                             READ FOR FUN:  ALMOST EVERY DAY
002 RD4FUN-2 (2          ) 1000                             READ FOR FUN:  ONCE OR TWICE A WEEK
003 RD4FUN-3 (3          ) 0100                             READ FOR FUN:  ONCE OR TWICE A MONTH
004 RD4FUN-4 (4          ) 0010                             READ FOR FUN:  NEVER OR HARDLY EVER
005 RD4FUN-? (M          ) 0001                             READ FOR FUN:  MISSING

CONDITIONING VARIABLE ID:  SUBJ0004
DESCRIPTION:               HOW OFTEN DO YOU TALK TO FRIENDS/FAMILY ABOUT WHAT YOU READ?
GRADES/ASSESSMENTS:        N04, S04, N08, N12
CONDITIONING VAR LABEL:    FAMLYRED
NAEP ID:                   R810902                          TOTAL NUMBER OF SPECIFIED CONTRASTS:     5
TYPE OF CONTRAST:          CLASS                            NUMBER OF INDEPENDENT CONTRASTS:         4

001 FAMRED-1 (1          ) 0000                             TALK ABOUT READING:  ALMOST EVERY DAY
002 FAMRED-2 (2          ) 1000                             TALK ABOUT READING:  ONCE OR TWICE A WEEK
003 FAMRED-3 (3          ) 0100                             TALK ABOUT READING:  ONCE OR TWICE A MONTH
004 FAMRED-4 (4          ) 0010                             TALK ABOUT READING:  NEVER OR HARDLY EVER
005 FAMRED-? (M          ) 0001                             TALK ABOUT READING:  MISSING

CONDITIONING VARIABLE ID:  SUBJ0005
DESCRIPTION:               HOW OFTEN DO YOU READ A STORIES OR NOVELS?
GRADES/ASSESSMENTS:        N08, N12
CONDITIONING VAR LABEL:    RD NOVEL
NAEP ID:                   R810904                          TOTAL NUMBER OF SPECIFIED CONTRASTS:     5
TYPE OF CONTRAST:          CLASS                            NUMBER OF INDEPENDENT CONTRASTS:         4

001 NOVEL-1  (1          ) 0000                             READ STORY/NOVEL:  ALMOST EVERY DAY
002 NOVEL-2  (2          ) 1000                             READ STORY/NOVEL:  ONCE OR TWICE A WEEK
003 NOVEL-3  (3          ) 0100                             READ STORY/NOVEL:  ONCE OR TWICE A MONTH
004 NOVEL-4  (4          ) 0010                             READ STORY/NOVEL:  NEVER OR HARDLY EVER
005 NOVEL-?  (M          ) 0001                             READ STORY/NOVEL:  MISSINGR HARDLY EVER
```

Table C-4 (continued)

Conditioning Variables Specific to the Main Reading Assessment

```
CONDITIONING VARIABLE ID:   SUBJ0006
DESCRIPTION:                HOW OFTEN DO YOU TAKE BOOKS OUT OF THE LIBRARY FOR YOUR OWN ENJOYMENT?
GRADES/ASSESSMENTS:         N04, S04, N08, N12
CONDITIONING VAR LABEL:     USELIBRY
NAEP ID:                    R810903                TOTAL NUMBER OF SPECIFIED CONTRASTS:    5
TYPE OF CONTRAST:           CLASS                  NUMBER OF INDEPENDENT CONTRASTS:        4

001 USELIB-1 (1           ) 0000                   USE THE LIBRARY:  ALMOST EVERY DAY
002 USELIB-2 (2           ) 1000                   USE THE LIBRARY:  ONCE OR TWICE A WEEK
003 USELIB-3 (3           ) 0100                   USE THE LIBRARY:  ONCE OR TWICE A MONTH
004 USELIB-4 (4           ) 0010                   USE THE LIBRARY:  NEVER OR HARDLY EVER
005 USELIB-? (M           ) 0001                   USE THE LIBRARY:  MISSINGR HARDLY EVER

CONDITIONING VARIABLE ID:   SUBJ0007
DESCRIPTION:                HOW OFTEN DO YOU READ A NEWSPAPER?
GRADES/ASSESSMENTS:         N08, N12
CONDITIONING VAR LABEL:     RD NEWSP
NAEP ID:                    R810905                TOTAL NUMBER OF SPECIFIED CONTRASTS:    5
TYPE OF CONTRAST:           CLASS                  NUMBER OF INDEPENDENT CONTRASTS:        4

001 RDNEWS-1 (1           ) 0000                   READ NEWSPAPER:  ALMOST EVERY DAY
002 RDNEWS-2 (2           ) 1000                   READ NEWSPAPER:  ONCE OR TWICE A WEEK
003 RDNEWS-3 (3           ) 0100                   READ NEWSPAPER:  ONCE OR TWICE A MONTH
004 RDNEWS-4 (4           ) 0010                   READ NEWSPAPER:  NEVER OR HARDLY EVER
005 RDNEWS-? (M           ) 0001                   READ NEWSPAPER:  MISSING

CONDITIONING VARIABLE ID:   SUBJ0008
DESCRIPTION:                HOW OFTEN DO YOU READ A MAGAZINE?
GRADES/ASSESSMENTS:         N08, N12
CONDITIONING VAR LABEL:     RD MAGAZ
NAEP ID:                    R810906                TOTAL NUMBER OF SPECIFIED CONTRASTS:    5
TYPE OF CONTRAST:           CLASS                  NUMBER OF INDEPENDENT CONTRASTS:        4

001 RDMAGZ-1 (1           ) 0000                   READ MAGAZINE:  ALMOST EVERY DAY
002 RDMAGZ-2 (2           ) 1000                   READ MAGAZINE:  ONCE OR TWICE A WEEK
003 RDMAGZ-3 (3           ) 0100                   READ MAGAZINE:  ONCE OR TWICE A MONTH
004 RDMAGZ-4 (4           ) 0010                   READ MAGAZINE:  NEVER OR HARDLY EVER
005 RDMAGZ-? (M           ) 0001                   READ MAGAZINE:  MISSING

CONDITIONING VARIABLE ID:   SUBJ0009
DESCRIPTION:                HOW OFTEN DOES YOUR TEACHER DISCUSS NEW OR DIFFICULT VOCABULARY?
GRADES/ASSESSMENTS:         N04, S04, N08, N12
CONDITIONING VAR LABEL:     VOCAB
NAEP ID:                    R811001                TOTAL NUMBER OF SPECIFIED CONTRASTS:    5
TYPE OF CONTRAST:           CLASS                  NUMBER OF INDEPENDENT CONTRASTS:        4

001 VOCAB-1  (1           ) 0000                   DISCUSS VOCABULARY:  ALMOST EVERY DAY
002 VOCAB-2  (2           ) 1000                   DISCUSS VOCABULARY:  ONCE OR TWICE A WEEK
003 VOCAB-3  (3           ) 0100                   DISCUSS VOCABULARY:  ONCE OR TWICE A MONTH
004 VOCAB-4  (4           ) 0010                   DISCUSS VOCABULARY:  NEVER OR HARDLY EVER
005 VOCAB-?  (M           ) 0001                   DISCUSS VOCABULARY:  MISSING

CONDITIONING VARIABLE ID:   SUBJ0010
DESCRIPTION:                HOW OFTEN DOES YOUR TEACHER ASK STUDENTS TO TALK TO EACH OTHER ABOUT WHAT THEY HAVE
                            READ?
GRADES/ASSESSMENTS:         N04, S04, N08, N12
CONDITIONING VAR LABEL:     TALKREAD
NAEP ID:                    R811002                TOTAL NUMBER OF SPECIFIED CONTRASTS:    5
TYPE OF CONTRAST:           CLASS                  NUMBER OF INDEPENDENT CONTRASTS:        4
```

Table C-4 (continued)

Conditioning Variables Specific to the Main Reading Assessment

```
001 TALKRD-1 (1          ) 0000                    TEACHER ASK TO TALK ABOUT READING:  ALMOST EVERY
                                                                                       DAY
002 TALKRD-2 (2          ) 1000                    TEACHER ASK TO TALK ABOUT READING:  ONCE OR TWICE
                                                                                       A WEEK
003 TALKRD-3 (3          ) 0100                    TEACHER ASK TO TALK ABOUT READING:  ONCE OR TWICE
                                                                                       A MONTH
004 TALKRD-4 (4          ) 0010                    TEACHER ASK TO TALK ABOUT READING:  NEVER OR
                                                                                       HARDLY EVER
005 TALKRD-? (M          ) 0001                    TEACHER ASK TO TALK ABOUT READING:  MISSING

CONDITIONING VARIABLE ID:  SUBJ0011
DESCRIPTION:               HOW OFTEN DOES YOUR TEACHER ASK YOU TO WORK IN A READING WORKBOOK OR ON A
                           WOKSHEET?
GRADES/ASSESSMENTS:        N04, S04, N08, N12
CONDITIONING VAR LABEL:    WBK/WSHT
NAEP ID:                   R811003                 TOTAL NUMBER OF SPECIFIED CONTRASTS:      5
TYPE OF CONTRAST:          CLASS                   NUMBER OF INDEPENDENT CONTRASTS:          4
001 WBK/WS-1 (1          ) 0000                    READING WORKBOOK/WORKSHEET:  ALMOST EVERY DAY
002 WBK/WS-2 (2          ) 1000                    READING WORKBOOK/WORKSHEET:  ONCE OR TWICE A WEEK
003 WBK/WS-3 (3          ) 0100                    READING WORKBOOK/WORKSHEET:  ONCE OR TWICE A MONTH
004 WBK/WS-4 (4          ) 0010                    READING WORKBOOK/WORKSHEET:  NEVER OR HARDLY EVER
005 WBK/WS-? (M          ) 0001                    READING WORKBOOK/WORKSHEET:  MISSING

CONDITIONING VARIABLE ID:  SUBJ0012
DESCRIPTION:               HOW OFTEN DOES YOUR TEACHER ASK YOU TO WRITE SOMETHING ABOUT WHAT YOU HAVE READ?
GRADES/ASSESSMENTS:        N04, S04, N08, N12
CONDITIONING VAR LABEL:    WRITREAD
NAEP ID:                   R811004                 TOTAL NUMBER OF SPECIFIED CONTRASTS:      5
TYPE OF CONTRAST:          CLASS                   NUMBER OF INDEPENDENT CONTRASTS:          4

001 WRTRED-1 (1          ) 0000                    WRITE ABOUT READING:  ALMOST EVERY DAY
002 WRTRED-2 (2          ) 1000                    WRITE ABOUT READING:  ONCE OR TWICE A WEEK
003 WRTRED-3 (3          ) 0100                    WRITE ABOUT READING:  ONCE OR TWICE A MONTH
004 WRTRED-4 (4          ) 0010                    WRITE ABOUT READING:  NEVER OR HARDLY EVER
005 WRTRED-? (M          ) 0001                    WRITE ABOUT READING:  MISSING

CONDITIONING VARIABLE ID:  SUBJ0013
DESCRIPTION:               HOW OFTEN DOES YOUR TEACHER ASK STUDENTS TO DO A GROUP ACTIVITY/PROJECT ABOUT WHAT
                           THEY HAVE READ?
GRADES/ASSESSMENTS:        N04, S04, N08, N12
CONDITIONING VAR LABEL:    READPROJ
NAEP ID:                   R811005                 TOTAL NUMBER OF SPECIFIED CONTRASTS:      5
TYPE OF CONTRAST:          CLASS                   NUMBER OF INDEPENDENT CONTRASTS:          4

001 RDPRJ-S1 (1          ) 0000                    PROJECT ABOUT READING: ALMOST EVERY DAY
002 RDPRJ-S2 (2          ) 1000                    PROJECT ABOUT READING: ONCE OR TWICE A WEEK
003 RDPRJ-S3 (3          ) 0100                    PROJECT ABOUT READING: ONCE OR TWICE A MONTH
004 RDPRJ-S4 (4          ) 0010                    PROJECT ABOUT READING: NEVER OR HARDLY EVER
005 RDPRJ-S? (M          ) 0001                    PROJECT ABOUT READING: MISSING

CONDITIONING VARIABLE ID:  SUBJ0014
DESCRIPTION:               HOW OFTEN DOES YOUR TEACHER ASK STUDENTS TO READ ALOUD?
GRADES/ASSESSMENTS:        N04, S04, N08, N12
CONDITIONING VAR LABEL:    RD ALOUD
NAEP ID:                   R811006                 TOTAL NUMBER OF SPECIFIED CONTRASTS:      5
TYPE OF CONTRAST:          CLASS                   NUMBER OF INDEPENDENT CONTRASTS:          4

001 ALOUD-S1 (1          ) 0000                    READ ALOUD: ALMOST EVERY DAY
002 ALOUD-S2 (2          ) 1000                    READ ALOUD: ONCE OR TWICE A WEEK
003 ALOUD-S3 (3          ) 0100                    READ ALOUD: ONCE OR TWICE A MONTH
004 ALOUD-S4 (4          ) 0010                    READ ALOUD: NEVER OR HARDLY EVER
005 ALOUD-S? (M          ) 0001                    READ ALOUD: MISSING
```

BEST COPY AVAILABLE

Table C-4 (continued)

Conditioning Variables Specific to the Main Reading Assessment

```
CONDITIONING VARIABLE ID:  SUBJ0015
DESCRIPTION:               HOW OFTEN DOES YOUR TEACHER ASK YOU TO READ SILENTLY?
GRADES/ASSESSMENTS:        N04, S04, N08, N12
CONDITIONING VAR LABEL:    RD SILNT
NAEP ID:                   R811007                  TOTAL NUMBER OF SPECIFIED CONTRASTS:    5
TYPE OF CONTRAST:          CLASS                    NUMBER OF INDEPENDENT CONTRASTS:        4

001 SILNT-S1 (1          ) 0000                     READ SILENTLY: ALMOST EVERY DAY
002 SILNT-S2 (2          ) 1000                     READ SILENTLY: ONCE OR TWICE A WEEK
003 SILNT-S3 (3          ) 0100                     READ SILENTLY: ONCE OR TWICE A MONTH
004 SILNT-S4 (4          ) 0010                     READ SILENTLY: NEVER OR HARDLY EVER
005 SILNT-S? (M          ) 0001                     READ SILENTLY: MISSING

CONDITIONING VARIABLE ID:  SUBJ0016
DESCRIPTION:               HOW OFTEN DOES YOUR TEACHER GIVE YOU TIME TO READ BOOKS YOU HAVE CHOSEN YOURSELF?
GRADES/ASSESSMENTS:        N04, S04, N08, N12
CONDITIONING VAR LABEL:    RDOWNBKS
NAEP ID:                   R811009                  TOTAL NUMBER OF SPECIFIED CONTRASTS:    5
TYPE OF CONTRAST:          CLASS                    NUMBER OF INDEPENDENT CONTRASTS:        4

001 OWNBK-S1 (1          ) 0000                     BOOKS CHOSEN YOURSELF:  ALMOST EVERY DAY
002 OWNBK-S2 (2          ) 1000                     BOOKS CHOSEN YOURSELF:  OR TWICE A WEEK
003 OWNBK-S3 (3          ) 0100                     BOOKS CHOSEN YOURSELF:  ONCE OR TWICE A MONTH
004 OWNBK-S4 (4          ) 0010                     BOOKS CHOSEN YOURSELF:  NEVER OR HARDLY EVER
005 OWNBK-S? (M          ) 0001                     BOOKS CHOSEN YOURSELF:  MISSING

CONDITIONING VARIABLE ID:  SUBJ0017
DESCRIPTION:               HOW OFTEN DOES YOUR TEACHER ASK YOU TO EXPLAIN OR SUPPORT YOUR UNDERSTANDING OF
                           WHAT YOU HAVE READ?
GRADES/ASSESSMENTS:        N08, N12
CONDITIONING VAR LABEL:    EXPLA RD
NAEP ID:                   R811010                  TOTAL NUMBER OF SPECIFIED CONTRASTS:    5
TYPE OF CONTRAST:          CLASS                    NUMBER OF INDEPENDENT CONTRASTS:        4

001 EXPLA-1  (1          ) 0000                     EXPLAIN WHAT YOU READ:  ALMOST EVERY DAY
002 EXPLA-2  (2          ) 1000                     EXPLAIN WHAT YOU READ:  OR TWICE A WEEK
003 EXPLA-3  (3          ) 0100                     EXPLAIN WHAT YOU READ:  ONCE OR TWICE A MONTH
004 EXPLA-4  (4          ) 0010                     EXPLAIN WHAT YOU READ:  NEVER OR HARDLY EVER
005 EXPLA-?  (M          ) 0001                     EXPLAIN WHAT YOU READ:  MISSING

CONDITIONING VARIABLE ID:  SUBJ0018
DESCRIPTION:               HOW OFTEN DOES YOUR TEACHER ASK YOU TO DISCUSS DIFFERENT INTERPRETATIONS OF WHAT
                           YOU HAVE READ?
GRADES/ASSESSMENTS:        N08, N12
CONDITIONING VAR LABEL:    DISCU RD
NAEP ID:                   R811011                  TOTAL NUMBER OF SPECIFIED CONTRASTS:    5
TYPE OF CONTRAST:          CLASS                    NUMBER OF INDEPENDENT CONTRASTS:        4

001 DISCU-1  (1          ) 0000                     DISCUSS INTERPRETATIONS OF READING:  ALMOST EVERY
                                                                                          DAY
002 DISCU-2  (2          ) 1000                     DISCUSS INTERPRETATIONS OF READING:  OR TWICE A
                                                                                          WEEK
003 DISCU-3  (3          ) 0100                     DISCUSS INTERPRETATIONS OF READING:  ONCE OR TWICE
                                                                                         A MONTH
004 DISCU-4  (4          ) 0010                     DISCUSS INTERPRETATIONS OF READING:  NEVER OR
                                                                                          HARDLY EVER
005 DISCU-?  (M          ) 0001                     DISCUSS INTERPRETATIONS OF READING:  MISSING

CONDITIONING VARIABLE ID:  SUBJ0019
DESCRIPTION:               DO YOU HAVE ACCESS TO A SCHOOL OR PUBLIC LIBRARY?
GRADES/ASSESSMENTS:        N08, N12
CONDITIONING VAR LABEL:    ACSS2LIB
NAEP ID:                   R820101                  TOTAL NUMBER OF SPECIFIED CONTRASTS:    3
TYPE OF CONTRAST:          CLASS                    NUMBER OF INDEPENDENT CONTRASTS:        2
```

Table C-4 (continued)

Conditioning Variables Specific to the Main Reading Assessment

```
001 AC2LIB-Y (1          ) 00                          ACCESS TO LIBRARY:  YES
002 AC2LIB-N (2          ) 10                          ACCESS TO LIBRARY:  NO
003 AC2LIB-? (M          ) 01                          ACCESS TO LIBRARY:  MISSING

CONDITIONING VARIABLE ID:  SUBJ0020
DESCRIPTION:               HOW OFTEN DO YOU USE A LIBRARY TO DO RESEARCH FOR A SCHOOL ASSIGNMENT?
GRADES/ASSESSMENTS:        N08, N12
CONDITIONING VAR LABEL:    LIB4RSCH
NAEP ID:                   R811301                     TOTAL NUMBER OF SPECIFIED CONTRASTS:     6
TYPE OF CONTRAST:          CLASS                       NUMBER OF INDEPENDENT CONTRASTS:         5

001 LIB4RS-1 (1          ) 00000                       LIBRARY FOR RESEARCH:  ALMOST EVERY DAY
002 LIB4RS-2 (2          ) 10000                       LIBRARY FOR RESEARCH:  ONCE OR TWICE A WEEK
003 LIB4RS-3 (3          ) 01000                       LIBRARY FOR RESEARCH:  ONCE OR TWICE A MONTH
004 LIB4RS-4 (4          ) 00100                       LIBRARY FOR RESEARCH:  ONCE OR TWICE A YEAR
005 LIB4RS-5 (5          ) 00010                       LIBRARY FOR RESEARCH:  NEVER OR HARDLY EVER
006 LIB4RS-? (M          ) 00001                       LIBRARY FOR RESEARCH:  MISSING

CONDITIONING VARIABLE ID:  SUBJ0021
DESCRIPTION:               HOW OFTEN DO YOU USE A LIBRARY TO BORROW BOOKS FOR A SCHOOL ASSIGNMENT?
GRADES/ASSESSMENTS:        N08, N12
CONDITIONING VAR LABEL:    LIB2BORW
NAEP ID:                   R811302                     TOTAL NUMBER OF SPECIFIED CONTRASTS:     6
TYPE OF CONTRAST:          CLASS                       NUMBER OF INDEPENDENT CONTRASTS:         5

001 LIB2BW-1 (1          ) 00000                       LIBRARY TO BORROW BOOKS:  ALMOST EVERY DAY
002 LIB2BW-2 (2          ) 10000                       LIBRARY TO BORROW BOOKS:  OR TWICE A WEEK
003 LIB2BW-3 (3          ) 01000                       LIBRARY TO BORROW BOOKS:  ONCE OR TWICE A MONTH
004 LIB2BW-4 (4          ) 00100                       LIBRARY TO BORROW BOOKS:  ONCE OR TWICE A YEAR
005 LIB2BW-5 (5          ) 00010                       LIBRARY TO BORROW BOOKS:  NEVER OR HARDLY EVER
006 LIB2BW-? (M          ) 00001                       LIBRARY TO BORROW BOOKS:  MISSING

CONDITIONING VARIABLE ID:  SUBJ0022
DESCRIPTION:               HOW OFTEN DO YOU USE A LIBRARY FOR INFORMATION FOR YOUR OWN USE?
GRADES/ASSESSMENTS:        N08, N12
CONDITIONING VAR LABEL:    LIB4YOU
NAEP ID:                   R811303                     TOTAL NUMBER OF SPECIFIED CONTRASTS:     6
TYPE OF CONTRAST:          CLASS                       NUMBER OF INDEPENDENT CONTRASTS:         5

001 LIB4U-1  (1          ) 00000                       LIBRARY FOR INFORMATION:  ALMOST EVERY DAY
002 LIB4U-2  (2          ) 10000                       LIBRARY FOR INFORMATION:  ONCE OR TWICE A WEEK
003 LIB4U-3  (3          ) 01000                       LIBRARY FOR INFORMATION:  ONCE OR TWICE A MONTH
004 LIB4U-4  (4          ) 00100                       LIBRARY FOR INFORMATION:  ONCE OR TWICE A YEAR
005 LIB4U-5  (5          ) 00010                       LIBRARY FOR INFORMATION:  NEVER OR HARDLY EVER
006 LIB4U-?  (M          ) 00001                       LIBRARY FOR INFORMATION:  MISSING

CONDITIONING VARIABLE ID:  SUBJ0023
DESCRIPTION:               HOW OFTEN DO YOU USE A LIBARY TO BORROW BOOKS/MATERIALS FOR YOU OWN ENJOYMENT?
GRADES/ASSESSMENTS:        N08, N12
CONDITIONING VAR LABEL:    LIB4NJOY
NAEP ID:                   R811305                     TOTAL NUMBER OF SPECIFIED CONTRASTS:     6
TYPE OF CONTRAST:          CLASS                       NUMBER OF INDEPENDENT CONTRASTS:         5

001 LIB4NJ-1 (1          ) 00000                       LIBRARY FOR ENJOYMENT:  ALMOST EVERY DAY
002 LIB4NJ-2 (2          ) 10000                       LIBRARY FOR ENJOYMENT:  ONCE OR TWICE A WEEK
003 LIB4NJ-3 (3          ) 01000                       LIBRARY FOR ENJOYMENT:  ONCE OR TWICE A MONTH
004 LIB4NJ-4 (4          ) 00100                       LIBRARY FOR ENJOYMENT:  ONCE OR TWICE A YEAR
005 LIB4NJ-5 (5          ) 00010                       LIBRARY FOR ENJOYMENT:  NEVER OR HARDLY EVER
006 LIB4NJ-? (M          ) 00001                       LIBRARY FOR ENJOYMENT:  MISSING

CONDITIONING VARIABLE ID:  SUBJ0024
DESCRIPTION:               HOW OFTEN DO YOU USE A LIBRARY AS A QUIET PLACE TO STUDY?
GRADES/ASSESSMENTS:        N08, N12
CONDITIONING VAR LABEL:    LIB4QIET
```

Table C-4 (continued)
Conditioning Variables Specific to the Main Reading Assessment

```
NAEP ID:                    R811304               TOTAL NUMBER OF SPECIFIED CONTRASTS:     6
TYPE OF CONTRAST:           CLASS                 NUMBER OF INDEPENDENT CONTRASTS:         5

001 LIB4QI-1 (1          ) 00000                  LIBRARY FOR QUIET STUDY:  ALMOST EVERY DAY
002 LIB4QI-2 (2          ) 10000                  LIBRARY FOR QUIET STUDY:  OR TWICE A WEEK
003 LIB4QI-3 (3          ) 01000                  LIBRARY FOR QUIET STUDY:  ONCE OR TWICE A MONTH
004 LIB4QI-4 (4          ) 00100                  LIBRARY FOR QUIET STUDY:  ONCE OR TWICE A YEAR,
                                                                            NEVER OR HARDLY EVER
005 LIB4QI-5 (5          ) 00010                  LIBRARY FOR QUIET STUDY:  ONCE OR TWICE A YEAR,
                                                                            NEVER OR HARDLY EVER
006 LIB4QI-? (M          ) 00001                  LIBRARY FOR QUIET STUDY:  MISSING

CONDITIONING VARIABLE ID:   SUBJ0025
DESCRIPTION:                HOW OFTEN DOES YOUR TEACHER TAKE YOU TO THE SCHOOL LIBRARY?
GRADES/ASSESSMENTS:         N04, S04
CONDITIONING VAR LABEL:     TAKE2LIB
NAEP ID:                    R811013               TOTAL NUMBER OF SPECIFIED CONTRASTS:     5
TYPE OF CONTRAST:           CLASS                 NUMBER OF INDEPENDENT CONTRASTS:         4

001 TK2LIB-1 (1          ) 0000                   TEACHER TAKE TO LIBRARY:  ALMOST EVERY DAY
002 TK2LIB-2 (2          ) 1000                   TEACHER TAKE TO LIBRARY:  ONCE OR TWICE A WEEK
003 TK2LIB-3 (3          ) 0100                   TEACHER TAKE TO LIBRARY:  ONCE OR TWICE A MONTH
004 TK2LIB-4 (4          ) 0010                   TEACHER TAKE TO LIBRARY:  NEVER OR HARDLY EVER
005 TK2LIB-? (M          ) 0001                   TEACHER TAKE TO LIBRARY:  MISSING

CONDITIONING VARIABLE ID:   SUBJ0026
DESCRIPTION:                DO YOU OR YOUR TEACHER SAVE YOUR READING WORK IN A FOLDER OR PORTFOLIO?
GRADES/ASSESSMENTS:         N04, S04, N08, N12
CONDITIONING VAR LABEL:     PRTFOLIO
NAEP ID:                    R820001               TOTAL NUMBER OF SPECIFIED CONTRASTS:     3
TYPE OF CONTRAST:           CLASS                 NUMBER OF INDEPENDENT CONTRASTS:         2

001 PFOLIO-Y (1          ) 00                     READING PORTFOLIO:  YES
002 PFOLIO-N (2          ) 10                     READING PORTFOLIO:  NO
003 PFOLIO-? (M          ) 01                     READING PORTFOLIO:  MISSING

CONDITIONING VARIABLE ID:   SUBJ0027
DESCRIPTION:                ARE YOU CURRENTLY ENROLLED IN OR HAVE YOU TAKEN AN AP COURSE IN ENGLISH?
GRADES/ASSESSMENTS:         N12
CONDITIONING VAR LABEL:     AP ENGLS
NAEP ID:                    R820201               TOTAL NUMBER OF SPECIFIED CONTRASTS:     3
TYPE OF CONTRAST:           CLASS                 NUMBER OF INDEPENDENT CONTRASTS:         2

001 AP ENG-Y (1          ) 00                     AP ENGLISH:  YES
002 AP ENG-N (2          ) 10                     AP ENGLISH:  NO
003 AP ENG-? (M          ) 01                     AP ENGLISH:  MISSING

CONDITIONING VARIABLE ID:   SUBJ0028
DESCRIPTION:                ABOUT HOW MANY QUESTIONS DID YOU GET RIGHT ON THE READING TEST?
GRADES/ASSESSMENTS:         N04, S04, N08, N12
CONDITIONING VAR LABEL:     #QUESTN+
NAEP ID:                    RN00101               TOTAL NUMBER OF SPECIFIED CONTRASTS:     5
TYPE OF CONTRAST:           CLASS                 NUMBER OF INDEPENDENT CONTRASTS:         4

001 #QUEST+1 (1          ) 0000                   NUMBER QUESTIONS RIGHT:  ALMOST ALL
002 #QUEST+2 (2          ) 1000                   NUMBER QUESTIONS RIGHT:  MORE THAN HALF
003 #QUEST+3 (3          ) 0100                   NUMBER QUESTIONS RIGHT:  ABOUT HALF
004 #QUEST+4 (4          ) 0010                   NUMBER QUESTIONS RIGHT:  LESS THAN HALF
005 #QUEST+? (M          ) 0001                   NUMBER QUESTIONS RIGHT:  MISSING

CONDITIONING VARIABLE ID:   SUBJ0029
DESCRIPTION:                HOW HARD WAS THIS READING TEST COMPARED TO OTHERS?
GRADES/ASSESSMENTS:         N04, S04, N08, N12
CONDITIONING VAR LABEL:     TEST DIF
```

Table C-4 (continued)

Conditioning Variables Specific to the Main Reading Assessment

```
NAEP ID:                    RM00201                      TOTAL NUMBER OF SPECIFIED CONTRASTS:     5
TYPE OF CONTRAST:           CLASS                        NUMBER OF INDEPENDENT CONTRASTS:         4

001 TESTDIF1 (1          ) 0000                          TEST DIFFICULTY:  MUCH HARDER THAN OTHERS
002 TESTDIF2 (2          ) 1000                          TEST DIFFICULTY:  HARDER THAN OTHERS
003 TESTDIF3 (3          ) 0100                          TEST DIFFICULTY:  ABOUT AS HARD AS OTHERS
004 TESTDIF4 (4          ) 0010                          TEST DIFFICULTY:  EASIER THAN OTHERS
005 TESTDIF? (M          ) 0001                          TEST DIFFICULTY:  MISSING

CONDITIONING VARIABLE ID:  SUBJ0030
DESCRIPTION:               HOW HARD DID YOU TRY ON THIS TEST COMPARED TO OTHER READING TESTS?
GRADES/ASSESSMENTS:        N04, S04, N08, N12
CONDITIONING VAR LABEL:    TEST EFF
NAEP ID:                   RM00301                       TOTAL NUMBER OF SPECIFIED CONTRASTS:     5
TYPE OF CONTRAST:          CLASS                         NUMBER OF INDEPENDENT CONTRASTS:         4

001 TESTEFF1 (1          ) 0000                          TEST EFFORT:  MUCH HARDER THAN OTHERS
002 TESTEFF2 (2          ) 1000                          TEST EFFORT:  HARDER THAN OTHERS
003 TESTEFF3 (3          ) 0100                          TEST EFFORT:  ABOUT AS HARD AS OTHERS
004 TESTEFF4 (4          ) 0010                          TEST EFFORT:  NOT AS HARD AS OTHERS
005 TESTEFF? (M          ) 0001                          TEST EFFORT:  MISSING

CONDITIONING VARIABLE ID:  SUBJ0031
DESCRIPTION:               HOW IMPORTANT WAS IT TO YOU TO DO WELL ON THE READING TEST?
GRADES/ASSESSMENTS:        N04, S04, N08, N12
CONDITIONING VAR LABEL:    TEST IMP
NAEP ID:                   RM00401                       TOTAL NUMBER OF SPECIFIED CONTRASTS:     5
TYPE OF CONTRAST:          CLASS                         NUMBER OF INDEPENDENT CONTRASTS:         4

001 TESTIMP1 (1          ) 0000                          TEST IMPORTANCE:  VERY IMPORTANT
002 TESTIMP2 (2          ) 1000                          TEST IMPORTANCE:  IMPORTANT
003 TESTIMP3 (3          ) 0100                          TEST IMPORTANCE:  SOMEWHAT IMPORTANT
004 TESTIMP4 (4          ) 0010                          TEST IMPORTANCE:  NOT VERY IMPORTANT
005 TESTIMP? (M          ) 0001                          TEST IMPORTANCE:  MISSING

CONDITIONING VARIABLE ID:  SUBJ0032
DESCRIPTION:               HOW OFTEN WERE YOU ASKED TO WRITE LONG ANSWERS ON READING TESTS?
GRADES/ASSESSMENTS:        N04, S04, N08, N12
CONDITIONING VAR LABEL:    LONG ANS
NAEP ID:                   RM00501                       TOTAL NUMBER OF SPECIFIED CONTRASTS:     5
TYPE OF CONTRAST:          CLASS                         NUMBER OF INDEPENDENT CONTRASTS:         4

001 LNGANSW1 (1          ) 0000                          LONG ANSWERS:  AT LEAST ONCE A WEEK
002 LNGANSW2 (2          ) 1000                          LONG ANSWERS:  ONCE OR TWICE A MONTH
003 LNGANSW3 (3          ) 0100                          LONG ANSWERS:  ONCE OR TWICE A YEAR
004 LNGANSW4 (4          ) 0010                          LONG ANSWERS:  NEVER
005 LNGANSW? (M          ) 0001                          LONG ANSWERS:  MISSING

CONDITIONING VARIABLE ID:  TCHR0001
DESCRIPTION:               TEACHER MATCH STATUS WITH STUDENT
GRADES/ASSESSMENTS:        N04, S04, N08
CONDITIONING VAR LABEL:    T_MATCH
NAEP ID:                   TCHMTCH                       TOTAL NUMBER OF SPECIFIED CONTRASTS:     3
TYPE OF CONTRAST:          CLASS                         NUMBER OF INDEPENDENT CONTRASTS:         2

001 TMCH-NO  (1,M        ) 00                            TEACHER MATCH:  NO MATCH
002 TMCH-PAR (2          ) 10                            TEACHER MATCH:  PARTIAL MATCH
003 TMCH-COM (3          ) 01                            TEACHER MATCH:  COMPLETE MATCH

CONDITIONING VARIABLE ID:  TCHR0002
DESCRIPTION:               TEACHER GENDER
GRADES/ASSESSMENTS:        N04, S04, N08
CONDITIONING VAR LABEL:    T_GENDER
NAEP ID:                   T040001                       TOTAL NUMBER OF SPECIFIED CONTRASTS:     3
```

Table C-4 (continued)

Conditioning Variables Specific to the Main Reading Assessment

```
TYPE OF CONTRAST:           CLASS                         NUMBER OF INDEPENDENT CONTRASTS:         2

001 T_MALE    (1          ) 00                            TEACHER GENDER:  MALE
002 T_FEMALE  (2          ) 10                            TEACHER GENDER:  FEMALE
003 T_SEX-?   (M          ) 01                            TEACHER GENDER:  MISSING, DOES NOT APPLY

CONDITIONING VARIABLE ID:   TCHR0003
DESCRIPTION:                YEARS TEACHING ELEMENTARY/SECONDARY SCHOOL
GRADES/ASSESSMENTS:         N04, S04, N08
CONDITIONING VAR LABEL:     T_YRSEXP
NAEP ID:                    T040301                       TOTAL NUMBER OF SPECIFIED CONTRASTS:     6
TYPE OF CONTRAST:           CLASS                         NUMBER OF INDEPENDENT CONTRASTS:         5

001 T_YREXP1  (1          ) 00000                         YEARS TEACHING:  2 OR LESS YEARS
002 T_YREXP2  (2          ) 10000                         YEARS TEACHING:  3-5 YEARS
003 T_YREXP3  (3          ) 01000                         YEARS TEACHING:  6-10 YEARS
004 T_YREXP4  (4          ) 00100                         YEARS TEACHING:  11-24 YEARS
005 T_YREXP5  (5          ) 00010                         YEARS TEACHING:  25 OR MORE YEARS
006 T_YREXP?  (M          ) 00001                         YEARS TEACHING:  MISSING, DOES NOT APPLY

CONDITIONING VARIABLE ID:   TCHR0004
DESCRIPTION:                TEACHER RACE/ETHNICITY
GRADES/ASSESSMENTS:         N04, S04, N08
CONDITIONING VAR LABEL:     T_RACE
NAEP ID:                    T050801                       TOTAL NUMBER OF SPECIFIED CONTRASTS:     7
TYPE OF CONTRAST:           CLASS                         NUMBER OF INDEPENDENT CONTRASTS:         6

001 T_WHITE   (1          ) 000000                        TEACHER RACE/ETHNICITY:  WHITE
002 T_BLACK   (2          ) 100000                        TEACHER RACE/ETHNICITY:  BLACK
003 T_HISP    (3          ) 010000                        TEACHER RACE/ETHNICITY:  HISPANIC
004 T_ASIAN   (4          ) 001000                        TEACHER RACE/ETHNICITY:  ASIAN
005 T_PAC IS  (5          ) 000100                        TEACHER RACE/ETHNICITY:  PACIFIC ISLANDER
006 T_AM IND  (6          ) 000010                        TEACHER RACE/ETHNICITY:  AMERICAN INDIAN/ALASKAN
                                                                                   NATIVE
007 T_RACE-?  (M          ) 000001                        TEACHER RACE/ETHNICITY:  MISSING, DOES NOT APPLY

CONDITIONING VARIABLE ID:   TCHR0005
DESCRIPTION:                TEACHER GENERAL CERTIFICATION (ELEMENTARY, MIDDLE/JUNIOR, HIGH SCHOOL EDUCATION)
GRADES/ASSESSMENTS:         N04, S04, N08
CONDITIONING VAR LABEL:     CERT GEN
NAEP ID:                    T040501                       TOTAL NUMBER OF SPECIFIED CONTRASTS:     4
TYPE OF CONTRAST:           CLASS                         NUMBER OF INDEPENDENT CONTRASTS:         3

001 CERTG-Y   (1          ) 000                           TEACHER GENERAL CERTIFICATION:  YES
002 CERTG-N   (2          ) 100                           TEACHER GENERAL CERTIFICATION:  NO
003 CERTG-NS  (3          ) 010                           TEACHER GENERAL CERTIFICATION:  NOT OFFERED IN
                                                                                          STATE
004 CERTG-?   (M          ) 001                           TEACHER GENERAL CERTIFICATION:  MISSING, DOES NOT
                                                                                          APPLY

CONDITIONING VARIABLE ID:   TCHR0006
DESCRIPTION:                TEACHER CERTIFICATION IN READING
GRADES/ASSESSMENTS:         N04, S04, N08
CONDITIONING VAR LABEL:     CERT RED
NAEP ID:                    T040502                       TOTAL NUMBER OF SPECIFIED CONTRASTS:     4
TYPE OF CONTRAST:           CLASS                         NUMBER OF INDEPENDENT CONTRASTS:         3

001 CERTR-Y   (1          ) 000                           TEACHER READING CERTIFICATION:  YES
002 CERTR-N   (2          ) 100                           TEACHER READING CERTIFICATION:  NO
003 CERTR-NS  (3          ) 010                           TEACHER READING CERTIFICATION:  NOT OFFERED IN
                                                                                          STATE
004 CERTR-?   (M          ) 001                           TEACHER READING CERTIFICATION:  MISSING, DOES NOT
                                                                                          APPLY
```

Table C-4 (continued)

Conditioning Variables Specific to the Main Reading Assessment

```
CONDITIONING VARIABLE ID:  TCHR0007
DESCRIPTION:               TEACHER CERTIFICATION MIDDLE/JUNIOR HIGH SCHOOL/SECONDARY ENGLISH/LANGUAGE ARTS
GRADES/ASSESSMENTS:        N04, S04, N08
CONDITIONING VAR LABEL:    CERT LAN
NAEP ID:                   T040508                          TOTAL NUMBER OF SPECIFIED CONTRASTS:     4
TYPE OF CONTRAST:          CLASS                            NUMBER OF INDEPENDENT CONTRASTS:         3

001 CERTL-Y  (1          ) 000                              TEACHER ENGLISH/LANGUAGE ARTS CERTIFICATION: YES
002 CERTL-N  (2          ) 100                              TEACHER ENGLISH/LANGUAGE ARTS CERTIFICATION: NO
003 CERTL-NS (3          ) 010                              TEACHER ENGLISH/LANGUAGE ARTS CERTIFICATION: NOT
                                                            OFFERED IN STATE
004 CERTL-?  (M          ) 001                              TEACHER ENGLISH/LANGUAGE ARTS CERTIFICATION:
                                                            MISSING, DOES NOT APPLY

CONDITIONING VARIABLE ID:  TCHR0008
DESCRIPTION:               TEACHER CERTIFICATION ENGLISH (8TH GRADE ONLY)
GRADES/ASSESSMENTS:        N08
CONDITIONING VAR LABEL:    CERT ENG
NAEP ID:                   T040511                          TOTAL NUMBER OF SPECIFIED CONTRASTS:     4
TYPE OF CONTRAST:          CLASS                            NUMBER OF INDEPENDENT CONTRASTS:         3

001 CERTE-Y  (1          ) 000                              TEACHER ENGLISH CERTIFICATION:  YES
002 CERTE-N  (2          ) 100                              TEACHER ENGLISH CERTIFICATION:  NO
003 CERTE-NS (3          ) 010                              TEACHER ENGLISH CERTIFICATION:  NOT OFFERED IN
                                                                                            STATE
004 CERTE-?  (M          ) 001                              TEACHER ENGLISH CERTIFICATION:  MISSING, DOES NOT
                                                                                            APPLY

CONDITIONING VARIABLE ID:  TCHR0009
DESCRIPTION:               HOW WELL DOES YOUR SCHOOL PROVIDE YOU WITH INSTRUCTIONAL MATERIAL/RESOURCES YOU
                           NEED?
GRADES/ASSESSMENTS:        N04, S04, N08
CONDITIONING VAR LABEL:    RESOURCE
NAEP ID:                   T041201                          TOTAL NUMBER OF SPECIFIED CONTRASTS:     5
TYPE OF CONTRAST:          CLASS                            NUMBER OF INDEPENDENT CONTRASTS:         4

001 RESOURC1 (1          ) 0000                             RESOURCES:  GET ALL
002 RESOURC2 (2          ) 1000                             RESOURCES:  GET MOST
003 RESOURC3 (3          ) 0100                             RESOURCES:  GET SOME
004 RESOURC4 (4          ) 0010                             RESOURCES:  DON'T GET
005 RESOURC? (M          ) 0001                             RESOURCES:  MISSING, DOES NOT APPLY

CONDITIONING VARIABLE ID:  TCHR0010
DESCRIPTION:               TEACHER UNDERGRADUATE MAJOR IN EDUCATION
GRADES/ASSESSMENTS:        N04, S04, N08
CONDITIONING VAR LABEL:    UGRAD ED
NAEP ID:                   T040701                          TOTAL NUMBER OF SPECIFIED CONTRASTS:     2
TYPE OF CONTRAST:          CLASS                            NUMBER OF INDEPENDENT CONTRASTS:         1

001 UGR ED-? (0,M        ) 0                                TEACHER UNDERGRAD EDUCATION MAJOR:  MISSING, DOES
                                                            NOT APPLY
002 UGR ED-Y (1          ) 1                                TEACHER UNDERGRAD EDUCATION MAJOR:  YES

CONDITIONING VARIABLE ID:  TCHR0011
DESCRIPTION:               TEACHER GRADUATE MAJOR IN EDUCATION
GRADES/ASSESSMENTS:        N04, S04, N08
CONDITIONING VAR LABEL:    GRAD ED
NAEP ID:                   T040801                          TOTAL NUMBER OF SPECIFIED CONTRASTS:     2
TYPE OF CONTRAST:          CLASS                            NUMBER OF INDEPENDENT CONTRASTS:         1

001 GRA ED-? (0,M        ) 0                                TEACHER GRADUATE EDUCATION MAJOR:  MISSING, DOES
                                                            NOT APPLY
002 GRA ED-Y (1          ) 1                                TEACHER GRADUATE EDUCATION MAJOR:  YES
```

Table C-4 (continued)

Conditioning Variables Specific to the Main Reading Assessment

```
CONDITIONING VARIABLE ID:  TCHR0012
DESCRIPTION:               NO TEACHER GRADUATE-LEVEL STUDY
GRADES/ASSESSMENTS:        N04, S04, N08
CONDITIONING VAR LABEL:    NO GRAD
NAEP ID:                   T040806                TOTAL NUMBER OF SPECIFIED CONTRASTS:     2
TYPE OF CONTRAST:          CLASS                  NUMBER OF INDEPENDENT CONTRASTS:         1

001 NOGRAD-? (0,M        ) 0                      NO TEACHER GRADUATE STUDY:  MISSING, DOES NOT
                                                  APPLY
002 NOGRAD-Y (1          ) 1                      NO TEACHER GRADUATE STUDY:  YES

CONDITIONING VARIABLE ID:  TCHR0013
DESCRIPTION:               HOW MANY YEARS IN TOTAL HAVE YOU TAUGHT READING? (4TH GRADE)
GRADES/ASSESSMENTS:        N04, S04
CONDITIONING VAR LABEL:    T4REDYRS
NAEP ID:                   T049901                TOTAL NUMBER OF SPECIFIED CONTRASTS:     6
TYPE OF CONTRAST:          CLASS                  NUMBER OF INDEPENDENT CONTRASTS:         5

001 T4REDYR1 (1          ) 00000                  YEARS TEACHING READING:  2 YEARS OR LESS
002 T4REDYR2 (2          ) 10000                  YEARS TEACHING READING:  3-5 YEARS
003 T4REDYR3 (3          ) 01000                  YEARS TEACHING READING:  6-10 YEARS
004 T4REDYR4 (4          ) 00100                  YEARS TEACHING READING:  11-24 YEARS
005 T4REDYR5 (5          ) 00010                  YEARS TEACHING READING:  25 YEARS OR MORE
006 T4REDYR? (M          ) 00001                  YEARS TEACHING READING:  MISSING, DOES NOT APPLY

CONDITIONING VARIABLE ID:  TCHR0014
DESCRIPTION:               HOW MANY YEARS IN TOTAL HAVE YOU TAUGHT READING? (8TH GRADE)
GRADES/ASSESSMENTS:        N08
CONDITIONING VAR LABEL:    T8REDYRS
NAEP ID:                   T050901                TOTAL NUMBER OF SPECIFIED CONTRASTS:     6
TYPE OF CONTRAST:          CLASS                  NUMBER OF INDEPENDENT CONTRASTS:         5

001 T8REDYR1 (1          ) 00000                  YEARS TEACHING READING:  2 YEARS OR LESS
002 T8REDYR2 (2          ) 10000                  YEARS TEACHING READING:  3-5 YEARS
003 T8REDYR3 (3          ) 01000                  YEARS TEACHING READING:  6-10 YEARS
004 T8REDYR4 (4          ) 00100                  YEARS TEACHING READING:  11-24 YEARS
005 T8REDYR5 (5          ) 00010                  YEARS TEACHING READING:  25 YEARS OR MORE
006 T8REDYR? (M          ) 00001                  YEARS TEACHING READING:  MISSING, DOES NOT APPLY

CONDITIONING VARIABLE ID:  TCHR0015
DESCRIPTION:               WHAT TYPE OF TEACHING CERTIFICATION DO YOU HAVE THAT IS RECOGNIZED BY THE STATE IN
                           WHICH YOU TEACH?
GRADES/ASSESSMENTS:        N04, S04, N08
CONDITIONING VAR LABEL:    TYP CERT
NAEP ID:                   T050001                TOTAL NUMBER OF SPECIFIED CONTRASTS:     5
TYPE OF CONTRAST:          CLASS                  NUMBER OF INDEPENDENT CONTRASTS:         4

001 CERT-NO  (1          ) 0000                   TYPE OF TEACHING CERTIFICATION:  NONE
002 CERT-TMP (2          ) 1000                   TYPE OF TEACHING CERTIFICATION:  TEMPORARY,
                                                  PROBATIONAL, PROVISIONAL, EMERGENCY
003 CERT-REG (3          ) 0100                   TYPE OF TEACHING CERTIFICATION:  REGULAR, BUT NOT
                                                  HIGHEST
004 CERT-HGH (4          ) 0010                   TYPE OF TEACHING CERTIFICATION:  HIGHEST AVAILABLE
005 CERT-?   (M          ) 0001                   TYPE OF TEACHING CERTIFICATION:  MISSING, DOES NOT
                                                  APPLY

CONDITIONING VARIABLE ID:  TCHR0016
DESCRIPTION:               WHAT IS THE HIGHEST ACADEMIC DEGREE YOU HOLD?
GRADES/ASSESSMENTS:        N04, S04, N08
CONDITIONING VAR LABEL:    T_DEGREE
NAEP ID:                   T050101                TOTAL NUMBER OF SPECIFIED CONTRASTS:     8
TYPE OF CONTRAST:          CLASS                  NUMBER OF INDEPENDENT CONTRASTS:         7
```

Table C-4 (continued)

Conditioning Variables Specific to the Main Reading Assessment

```
001 TDEG-HSD (1          ) 0000000              TEACHER HIGHEST DEGREE:  HIGH SCHOOL DIPLOMA
002 TDEG-ASC (2          ) 1000000              TEACHER HIGHEST DEGREE:  ASSOCIATES/VOCATIONAL
003 TDEG-BAC (3          ) 0100000              TEACHER HIGHEST DEGREE:  BACHELOR'S HIGHEST DEGREE
004 TDEG-MAS (4          ) 0010000              TEACHER HIGHEST DEGREE:  MASTER'S HIGHEST DEGREE
005 TDEG-EDS (5          ) 0001000              TEACHER HIGHEST DEGREE:  EDUCATION SPECIALIST
006 TDEG-DOC (6          ) 0000100              TEACHER HIGHEST DEGREE:  DOCTORATE
007 TDEG-PRO (7          ) 0000010              TEACHER HIGHEST DEGREE:  PROFESSIONAL HIGHEST
                                                                         DEGREE
008 TDEG-?   (M          ) 0000001              TEACHER HIGHEST DEGREE:  MISSING, DOES NOT APPLY

CONDITIONING VARIABLE ID:  TCHR0017
DESCRIPTION:               TEACHER ENGLISH UNDERGRADUATE MAJOR
GRADES/ASSESSMENTS:        N04, S04, N08
CONDITIONING VAR LABEL:    UGRD ENG
NAEP ID:                   T040706              TOTAL NUMBER OF SPECIFIED CONTRASTS:     2
TYPE OF CONTRAST:          CLASS                NUMBER OF INDEPENDENT CONTRASTS:         1

001 UGRENG-Y (1          ) 0                    TEACHER ENGLISH UNDERGRADUATE MAJOR:  YES
002 UGRENG-? (0,M        ) 1                    TEACHER ENGLISH UNDERGRADUATE MAJOR:  MISSING,
                                                DOES NOT APPLY

CONDITIONING VARIABLE ID:  TCHR0018
DESCRIPTION:               TEACHER READING/LANGUAGE ARTS UNDERGRADUATE MAJOR
GRADES/ASSESSMENTS:        N04, S04, N08
CONDITIONING VAR LABEL:    UGRD RED
NAEP ID:                   T040707              TOTAL NUMBER OF SPECIFIED CONTRASTS:     2
TYPE OF CONTRAST:          CLASS                NUMBER OF INDEPENDENT CONTRASTS:         1

001 UGRRED-Y (1          ) 0                    TEACHER READING UNDERGRADUATE MAJOR:  YES
002 UGRRED-? (0,M        ) 1                    TEACHER READING UNDERGRADUATE MAJOR:  MISSING,
                                                DOES NOT APPLY

CONDITIONING VARIABLE ID:  TCHR0019
DESCRIPTION:               TEACHER ENGLISH GRADUATE MAJOR
GRADES/ASSESSMENTS:        N04, S04, N08
CONDITIONING VAR LABEL:    GRAD ENG
NAEP ID:                   T040807              TOTAL NUMBER OF SPECIFIED CONTRASTS:     2
TYPE OF CONTRAST:          CLASS                NUMBER OF INDEPENDENT CONTRASTS:         1

001 GRDENG-Y (1          ) 0                    TEACHER ENGLISH GRADUATE MAJOR: YES
002 GRDENG-? (0,M        ) 1                    TEACHER ENGLISH GRADUATE MAJOR: MISSING,
                                                                                DOES NOT APPLY

CONDITIONING VARIABLE ID:  TCHR0020
DESCRIPTION:               TEACHER READING/LANGUAGE ARTS GRADUATE MAJOR
GRADES/ASSESSMENTS:        N04, S04, N08
CONDITIONING VAR LABEL:    GRAD RED
NAEP ID:                   T040808              TOTAL NUMBER OF SPECIFIED CONTRASTS:     2
TYPE OF CONTRAST:          CLASS                NUMBER OF INDEPENDENT CONTRASTS:         1

001 GRDRED-Y (1          ) 0                    TEACHER READING GRADUATE MAJOR: YES
002 GRDRED-? (0,M        ) 1                    TEACHER READING GRADUATE MAJOR: MISSING,
                                                                                DOES NOT APPLY

CONDITIONING VARIABLE ID:  TCHR0021
DESCRIPTION:               HOW MUCH TIME HAVE YOU SPENT LAST YEAR IN READING DEVELOPMENT WORKSHOPS/SEMINARS?
GRADES/ASSESSMENTS:        N04, S04, N08
CONDITIONING VAR LABEL:    READ DEV
NAEP ID:                   T050201              TOTAL NUMBER OF SPECIFIED CONTRASTS:     6
TYPE OF CONTRAST:          CLASS                NUMBER OF INDEPENDENT CONTRASTS:         5

001 READDEV1 (1          ) 00000                TIME IN READING DEVELOPMENT WORKSHOPS:  NONE
002 READDEV2 (2          ) 10000                TIME IN READING DEVELOPMENT WORKSHOPS:  LESS THAN
                                                                                        6 HOURS
```

BEST COPY AVAILABLE

Table C-4 (continued)

Conditioning Variables Specific to the Main Reading Assessment

```
003 READDEV3 (3          ) 01000                 TIME IN READING DEVELOPMENT WORKSHOPS:  6-15 HOURS
004 READDEV4 (4          ) 00100                 TIME IN READING DEVELOPMENT WORKSHOPS:  16-35
                                                 HOURS
005 READDEV5 (5          ) 00010                 TIME IN READING DEVELOPMENT WORKSHOPS:  MORE THAN
                                                 35 HOURS
006 READDEV5 (M          ) 00001                 TIME IN READING DEVELOPMENT WORKSHOPS:  MISSING,
                                                 DOES NOT APPLY

CONDITIONING VARIABLE ID:  TCHR0022
DESCRIPTION:               HOW MANY HOURS DO YOU HAVE DESIGNATED AS PREPERATION PERIODS PER WEEK?
GRADES/ASSESSMENTS:        N04, S04, N08
CONDITIONING VAR LABEL:    PREP PER
NAEP ID:                   T051101               TOTAL NUMBER OF SPECIFIED CONTRASTS:      5
TYPE OF CONTRAST:          CLASS                 NUMBER OF INDEPENDENT CONTRASTS:          4

001 PRPER-0  (1          ) 0000                  TEACHER WEEKLY PREPARTION PERIODS:  NONE
002 PRPER-<1 (2          ) 1000                  TEACHER WEEKLY PREPARTION PERIODS:  LESS THAN 1
003 PRPER-12 (3          ) 0100                  TEACHER WEEKLY PREPARTION PERIODS:  1 TO 2
004 PRPER->2 (4          ) 0010                  TEACHER WEEKLY PREPARTION PERIODS:  MORE THAN 2
005 PRPER-?  (M          ) 0001                  TEACHER WEEKLY PREPARTION PERIODS:  MISSING, DOES
                                                 NOT APPLY

CONDITIONING VARIABLE ID:  TCHR0023
DESCRIPTION:               ARE CURRICULUM SPECIALISTS AVAILABLE FOR READING?
GRADES/ASSESSMENTS:        N04, S04, N08
CONDITIONING VAR LABEL:    CURSPECS
NAEP ID:                   T041301               TOTAL NUMBER OF SPECIFIED CONTRASTS:      3
TYPE OF CONTRAST:          CLASS                 NUMBER OF INDEPENDENT CONTRASTS:          2

001 CSPECS-Y (1          ) 00                    READING CURRICULUM SPECIALISTS:  YES
002 CSPECS-N (2          ) 10                    READING CURRICULUM SPECIALISTS:  NO
003 CSPECS-? (M          ) 01                    READING CURRICULUM SPECIALISTS:  MISSING, DOES NOT
                                                 APPLY

CONDITIONING VARIABLE ID:  TSUB0001
DESCRIPTION:               WHAT IS YOUR AVERAGE READING CLASS SIZE? (4TH GRADE)
GRADES/ASSESSMENTS:        N04, S04
CONDITIONING VAR LABEL:    CLASSIZ4
NAEP ID:                   T050701               TOTAL NUMBER OF SPECIFIED CONTRASTS:      6
TYPE OF CONTRAST:          CLASS                 NUMBER OF INDEPENDENT CONTRASTS:          5

001 CLASIZ-1 (1          ) 00000                 AVERAGE READING CLASS SIZE:  1-20 STUDENTS
002 CLASIZ-2 (2          ) 10000                 AVERAGE READING CLASS SIZE:  21-25 STUDENTS
003 CLASIZ-3 (3          ) 01000                 AVERAGE READING CLASS SIZE:  26-30 STUDENTS
004 CLASIZ-4 (4          ) 00100                 AVERAGE READING CLASS SIZE:  31-35 STUDENTS
005 CLASIZ-5 (5          ) 00010                 AVERAGE READING CLASS SIZE:  36 OR MORE STUDENTS
006 CLASIZ-? (M          ) 00001                 AVERAGE READING CLASS SIZE:  MISSING, DOES NOT
                                                 APPLY

CONDITIONING VARIABLE ID:  TSUB0002
DESCRIPTION:               WHAT IS THE NUMBER OF STUDENTS IN EACH READING CLASS? (8TH GRADE)
GRADES/ASSESSMENTS:        N08
CONDITIONING VAR LABEL:    CLASSIZ8
NAEP ID:                   T051001               TOTAL NUMBER OF SPECIFIED CONTRASTS:      6
TYPE OF CONTRAST:          CLASS                 NUMBER OF INDEPENDENT CONTRASTS:          5

001 CLASIZ-1 (1          ) 00000                 AVERAGE READING CLASS SIZE:  1-20 STUDENTS
002 CLASIZ-2 (2          ) 10000                 AVERAGE READING CLASS SIZE:  21-25 STUDENTS
003 CLASIZ-3 (3          ) 01000                 AVERAGE READING CLASS SIZE:  26-30 STUDENTS
004 CLASIZ-4 (4          ) 00100                 AVERAGE READING CLASS SIZE:  31-35 STUDENTS
005 CLASIZ-5 (5          ) 00010                 AVERAGE READING CLASS SIZE:  36 OR MORE STUDENTS
006 CLASIZ-? (M          ) 00001                 AVERAGE READING CLASS SIZE:  MISSING, DOES NOT
                                                 APPLY
```

Table C-4 (continued)

Conditioning Variables Specific to the Main Reading Assessment

```
CONDITIONING VARIABLE ID:  TSUB0003
DESCRIPTION:               ARE STUDENTS ASSIGNED TO THIS READING CLASS BY ABILITY? (TEACHER)
GRADES/ASSESSMENTS:        N04, S04, N08
CONDITIONING VAR LABEL:    T_ABILTY
NAEP ID:                   T046101                  TOTAL NUMBER OF SPECIFIED CONTRASTS:     3
TYPE OF CONTRAST:          CLASS                    NUMBER OF INDEPENDENT CONTRASTS:         2

001 T_ABIL-Y (1          ) 00                      STUDENTS ASSIGNED TO READING CLASS BY ABILITY:  YES
002 T_ABIL-N (2          ) 10                      STUDENTS ASSIGNED TO READING CLASS BY ABILITY:  NO
003 T_ABIL-? (M          ) 01                      STUDENTS ASSIGNED TO READING CLASS BY ABILITY:
                                                   MISSING, DOES NOT APPLY

CONDITIONING VARIABLE ID:  TSUB0004
DESCRIPTION:               WHAT IS THE READING ABILITY LEVEL OF STUDENTS IN THIS CLASS?
GRADES/ASSESSMENTS:        N04, S04, N08
CONDITIONING VAR LABEL:    ABIL RED
NAEP ID:                   T046201                  TOTAL NUMBER OF SPECIFIED CONTRASTS:     5
TYPE OF CONTRAST:          CLASS                    NUMBER OF INDEPENDENT CONTRASTS:         4

001 ABILRED1 (1          ) 0000                     READING ABILITY:  PRIMARILY HIGH
002 ABILRED2 (2          ) 1000                     READING ABILITY:  PRIMARILY AVERAGE
003 ABILRED3 (3          ) 0100                     READING ABILITY:  PRIMARILY LOW
004 ABILRED4 (4          ) 0010                     READING ABILITY:  WIDELY MIXED
005 ABILRED? (M          ) 0001                     READING ABILITY:  MISSING, DOES NOT APPLY

CONDITIONING VARIABLE ID:  TSUB0005
DESCRIPTION:               HOW MUCH TIME DO YOU SPEND WITH THIS CLASS FOR READING INSTRUCTION EACH DAY?
GRADES/ASSESSMENTS:        N04, S04, N08
CONDITIONING VAR LABEL:    INSTTIME
NAEP ID:                   T046301                  TOTAL NUMBER OF SPECIFIED CONTRASTS:     5
TYPE OF CONTRAST:          CLASS                    NUMBER OF INDEPENDENT CONTRASTS:         4

001 INSTIME1 (1          ) 0000                     READING INSTRUCTION TIME:  30 MINUTES A DAY
002 INSTIME2 (2          ) 1000                     READING INSTRUCTION TIME:  45 MINUTES A DAY
003 INSTIME3 (3          ) 0100                     READING INSTRUCTION TIME:  60 MINUTES A DAY
004 INSTIME4 (4          ) 0010                     READING INSTRUCTION TIME:  90 MINUTES A DAY
005 INSTIME? (M          ) 0001                     READING INSTRUCTION TIME:  MISSING, DOES NOT APPLY

CONDITIONING VARIABLE ID:  TSUB0006
DESCRIPTION:               HOW MANY INSTRUCTIONAL GROUPS DO YOU DIVIDE YOUR READING CLASS?
GRADES/ASSESSMENTS:        N04, S04, N08
CONDITIONING VAR LABEL:    INSTGRPS
NAEP ID:                   T046401                  TOTAL NUMBER OF SPECIFIED CONTRASTS:     8
TYPE OF CONTRAST:          CLASS                    NUMBER OF INDEPENDENT CONTRASTS:         7

001 1 GROUP  (1          ) 0000000                  NUMBER INSTRUCTIONAL GROUPS:  WHOLE-CLASS ACTIVITY
002 FLEX GRP (2          ) 1000000                  NUMBER INSTRUCTIONAL GROUPS:  FLEXIBLE GROUPING
003 2 GROUPS (3          ) 0100000                  NUMBER INSTRUCTIONAL GROUPS:  2 GROUPS
004 3 GROUPS (4          ) 0010000                  NUMBER INSTRUCTIONAL GROUPS:  3 GROUPS
005 4 GROUPS (5          ) 0001000                  NUMBER INSTRUCTIONAL GROUPS:  4 GROUPS
006 5+GROUPS (6          ) 0000100                  NUMBER INSTRUCTIONAL GROUPS:  5 OR MORE GROUPS
007 INDIVLZD (7          ) 0000010                  NUMBER INSTRUCTIONAL GROUPS:  INDIVIDUALIZED INSTR
008 GROUPS-? (M          ) 0000001                  NUMBER INSTRUCTIONAL GROUPS:  MISSING, DOES NOT
                                                                                  APPLY

CONDITIONING VARIABLE ID:  TSUB0007
DESCRIPTION:               WHAT TYPE OF MATERIALS FORM THE CORE OF YOUR READING PROGRAM?
GRADES/ASSESSMENTS:        N04, S04, N08
CONDITIONING VAR LABEL:    READMATS
NAEP ID:                   T046501                  TOTAL NUMBER OF SPECIFIED CONTRASTS:     5
TYPE OF CONTRAST:          CLASS                    NUMBER OF INDEPENDENT CONTRASTS:         4
```

Table C-4 (continued)

Conditioning Variables Specific to the Main Reading Assessment

```
001 BASAL    (1          ) 0000                          TYPE OF READING MATERIALS:  BASAL
002 TRADE    (2          ) 1000                          TYPE OF READING MATERIALS:  TRADE
003 BAS/TRAD (3          ) 0100                          TYPE OF READING MATERIALS:  BASAL AND TRADE
004 OTHER RM (4          ) 0010                          TYPE OF READING MATERIALS:  OTHER
005 RDMATS-? (M          ) 0001                          TYPE OF READING MATERIALS:  MISSING, DOES NOT
                                                         APPLY

CONDITIONING VARIABLE ID:  TSUB0008
DESCRIPTION:               HOW OFTEN ARE CHILDREN'S NEWSPAPERS/MAGAZINES USED IN READING CLASS?
GRADES/ASSESSMENTS:        N04, S04
CONDITIONING VAR LABEL:    CHILDMAG
NAEP ID:                   T046601                       TOTAL NUMBER OF SPECIFIED CONTRASTS:      5
TYPE OF CONTRAST:          CLASS                         NUMBER OF INDEPENDENT CONTRASTS:          4

001 CHIMAG-1 (1          ) 0000                          NEWSPAPERS/MAGAZINES (TEACHER):  ALMOST EVERY DAY
002 CHIMAG-2 (2          ) 1000                          NEWSPAPERS/MAGAZINES (TEACHER):  ONCE OR TWICE A
                                                         WEEK
003 CHIMAG-3 (3          ) 0100                          NEWSPAPERS/MAGAZINES (TEACHER):  ONCE OR TWICE A
                                                         MONTH
004 CHIMAG-4 (4          ) 0010                          NEWSPAPERS/MAGAZINES (TEACHER):  NEVER OF HARDLEY
                                                         EVER
005 CHIMAG-? (M          ) 0001                          NEWSPAPERS/MAGAZINES (TEACHER):  MISSING, DOES NOT
                                                         APPLY

CONDITIONING VARIABLE ID:  TSUB0009
DESCRIPTION:               HOW OFTEN ARE READING KITS USED IN READING CLASS?
GRADES/ASSESSMENTS:        N04, S04, N08
CONDITIONING VAR LABEL:    READKITS
NAEP ID:                   T046602                       TOTAL NUMBER OF SPECIFIED CONTRASTS:      5
TYPE OF CONTRAST:          CLASS                         NUMBER OF INDEPENDENT CONTRASTS:          4

001 RDKITS-1 (1          ) 0000                          READING KITS (TEACHER):  ALMOST EVERY DAY
002 RDKITS-1 (2          ) 1000                          READING KITS (TEACHER):  ONCE OR TWICE A WEEK
003 RDKITS-2 (3          ) 0100                          READING KITS (TEACHER):  ONCE OR TWICE A MONTH
004 RDKITS-3 (4          ) 0010                          READING KITS (TEACHER):  NEVER OF HARDLEY EVER
005 RDKITS-? (M          ) 0001                          READING KITS (TEACHER):  MISSING, DOES NOT APPLY

CONDITIONING VARIABLE ID:  TSUB0010
DESCRIPTION:               HOW OFTEN IS READING COMPUTER SOFTWARE IN READING CLASS?
GRADES/ASSESSMENTS:        N04, S04, N08
CONDITIONING VAR LABEL:    SOFTWARE
NAEP ID:                   T046603                       TOTAL NUMBER OF SPECIFIED CONTRASTS:      5
TYPE OF CONTRAST:          CLASS                         NUMBER OF INDEPENDENT CONTRASTS:          4

001 SOFTWR-1 (1          ) 0000                          READING COMPUTER SOFTWARE (TEACHER):  ALMOST EVERY
                                                         DAY
002 SOFTWR-2 (2          ) 1000                          READING COMPUTER SOFTWARE (TEACHER):  ONCE OR
                                                         TWICE A WEEK
003 SOFTWR-3 (3          ) 0100                          READING COMPUTER SOFTWARE (TEACHER):  ONCE OR
                                                         TWICE A MONTH
004 SOFTWR-4 (4          ) 0010                          READING COMPUTER SOFTWARE (TEACHER):  NEVER OF
                                                         HARDLEY EVER
005 SOFTWR-? (M          ) 0001                          READING COMPUTER SOFTWARE (TEACHER):  MISSING,
                                                         DOES NOT APPLY

CONDITIONING VARIABLE ID:  TSUB0011
DESCRIPTION:               HOW OFTEN ARE VARIETY OF BOOKS USED IN READING CLASS?
GRADES/ASSESSMENTS:        N04, S04, N08
CONDITIONING VAR LABEL:    VARTYBKS
NAEP ID:                   T046604                       TOTAL NUMBER OF SPECIFIED CONTRASTS:      5
TYPE OF CONTRAST:          CLASS                         NUMBER OF INDEPENDENT CONTRASTS:          4
```

Table C-4 (continued)

Conditioning Variables Specific to the Main Reading Assessment

```
001 VARBKS-1 (1           ) 0000                      VARIETY OF BOOKS (TEACHER):  ALMOST EVERY DAY
002 VARBKS-2 (2           ) 1000                      VARIETY OF BOOKS (TEACHER):  ONCE OR TWICE A WEEK
003 VARBKS-3 (3           ) 0100                      VARIETY OF BOOKS (TEACHER):  ONCE OR TWICE A MONTH
004 VARBKS-4 (4           ) 0010                      VARIETY OF BOOKS (TEACHER):  NEVER OF HARDLEY EVER
005 VARBKS-? (M           ) 0001                      VARIETY OF BOOKS (TEACHER):  MISSING, DOES NOT
                                                      APPLY

CONDITIONING VARIABLE ID:  TSUB0012
DESCRIPTION:               HOW OFTEN ARE MATERIALS FROM OTHER SUBJECTS USED IN READING CLASS?
GRADES/ASSESSMENTS:        N04, S04, N08
CONDITIONING VAR LABEL:    OTHRMATS
NAEP ID:                   T046605                    TOTAL NUMBER OF SPECIFIED CONTRASTS:      5
TYPE OF CONTRAST:          CLASS                      NUMBER OF INDEPENDENT CONTRASTS:          4

001 OTHMAT-1 (1           ) 0000                      OTHER SUBJECT MATERIALS (TEACHER):  ALMOST EVERY
                                                      DAY
002 OTHMAT-2 (2           ) 1000                      OTHER SUBJECT MATERIALS (TEACHER):  ONCE OR TWICE
                                                      A WEEK
003 OTHMAT-3 (3           ) 0100                      OTHER SUBJECT MATERIALS (TEACHER):  ONCE OR TWICE
                                                      A MONTH
004 OTHMAT-4 (4           ) 0010                      OTHER SUBJECT MATERIALS (TEACHER):  NEVER OR
                                                      HARDLY EVER
005 OTHMAT-? (M           ) 0001                      OTHER SUBJECT MATERIALS (TEACHER):  MISSING, DOES
                                                      NOT APPLY

CONDITIONING VARIABLE ID:  TSUB0013
DESCRIPTION:               HOW OFTEN DO YOU DISCUSS NEW OR DIFFICULT VOCABULARY?
GRADES/ASSESSMENTS:        N04, S04, N08
CONDITIONING VAR LABEL:    T_VOCAB
NAEP ID:                   T046701                    TOTAL NUMBER OF SPECIFIED CONTRASTS:      5
TYPE OF CONTRAST:          CLASS                      NUMBER OF INDEPENDENT CONTRASTS:          4

001 VOCAB-T1 (1           ) 0000                      DISCUSS VOCABULARY:  ALMOST EVERY DAY
002 VOCAB-T2 (2           ) 1000                      DISCUSS VOCABULARY:  ONCE OR TWICE A WEEK
003 VOCAB-T3 (3           ) 0100                      DISCUSS VOCABULARY:  ONCE OR TWICE A MONTH
004 VOCAB-T4 (4           ) 0010                      DISCUSS VOCABULARY:  NEVER OR HARDLY EVER
005 VOCAB-T? (M           ) 0001                      DISCUSS VOCABULARY:  MISSING, DOES NOT APPLY

CONDITIONING VARIABLE ID:  TSUB0014
DESCRIPTION:               HOW OFTEN DO YOU ASK STUDENTS TO READ ALOUD?
GRADES/ASSESSMENTS:        N04, S04, N08
CONDITIONING VAR LABEL:    T_ALOUD
NAEP ID:                   T046702                    TOTAL NUMBER OF SPECIFIED CONTRASTS:      5
TYPE OF CONTRAST:          CLASS                      NUMBER OF INDEPENDENT CONTRASTS:          4

001 ALOUD-T1 (1           ) 0000                      READ ALOUD: ALMOST EVERY DAY
002 ALOUD-T2 (2           ) 1000                      READ ALOUD: ONCE OR TWICE A WEEK
003 ALOUD-T3 (3           ) 0100                      READ ALOUD: ONCE OR TWICE A MONTH
004 ALOUD-T4 (4           ) 0010                      READ ALOUD: NEVER OR HARDLY EVER
005 ALOUD-T? (M           ) 0001                      READ ALOUD: MISSING, DOES NOT APPLY

CONDITIONING VARIABLE ID:  TSUB0015
DESCRIPTION:               HOW OFTEN DO YOU ASK STUDENTS TO TALK TO EACH OTHER ABOUT WHAT THEY HAVE READ?
GRADES/ASSESSMENTS:        N04, S04, N08
CONDITIONING VAR LABEL:    T_TALKRD
NAEP ID:                   T046703                    TOTAL NUMBER OF SPECIFIED CONTRASTS:      5
TYPE OF CONTRAST:          CLASS                      NUMBER OF INDEPENDENT CONTRASTS:          4

001 TLKRD-T1 (1           ) 0000                      TALK ABOUT READING:  ALMOST EVERY DAY
002 TLKRD-T2 (2           ) 1000                      TALK ABOUT READING:  ONCE OR TWICE A WEEK
003 TLKRD-T3 (3           ) 0100                      TALK ABOUT READING:  ONCE OR TWICE A MONTH
004 TLKRD-T4 (4           ) 0010                      TALK ABOUT READING:  NEVER OR HARDLY EVER
005 TLKRD-T? (M           ) 0001                      TALK ABOUT READING:  MISSING, DOES NOT APPLY
```

Table C-4 (continued)

Conditioning Variables Specific to the Main Reading Assessment

```
CONDITIONING VARIABLE ID:  TSUB0016
DESCRIPTION:               HOW OFTEN DO YOU ASK STUDENTS TO WRITE SOMETHING ABOUT WHAT THEY HAVE READ?
GRADES/ASSESSMENTS:        N04, S04, N08
CONDITIONING VAR LABEL:    T_WRITRD
NAEP ID:                   T046704                 TOTAL NUMBER OF SPECIFIED CONTRASTS:    5
TYPE OF CONTRAST:          CLASS                   NUMBER OF INDEPENDENT CONTRASTS:        4

001 WRTRD-T1 (1          ) 0000                    WRITE ABOUT READING:  ALMOST EVERY DAY
002 WRTRD-T2 (2          ) 1000                    WRITE ABOUT READING:  ONCE OR TWICE A WEEK
003 WRTRD-T3 (3          ) 0100                    WRITE ABOUT READING:  ONCE OR TWICE A MONTH
004 WRTRD-T4 (4          ) 0010                    WRITE ABOUT READING:  NEVER OR HARDLY EVER
005 WRTRD-T? (M          ) 0001                    WRITE ABOUT READING:  MISSING, DOES NOT APPLY

CONDITIONING VARIABLE ID:  TSUB0017
DESCRIPTION:               HOW OFTEN DO YOU ASK STUDENTS TO WORK IN A READING WORKBOOK OR ON A WORKSHEET?
GRADES/ASSESSMENTS:        N04, S04, N08
CONDITIONING VAR LABEL:    T_WBKWSH
NAEP ID:                   T046705                 TOTAL NUMBER OF SPECIFIED CONTRASTS:    5
TYPE OF CONTRAST:          CLASS                   NUMBER OF INDEPENDENT CONTRASTS:        4

001 WB/WS-T1 (1          ) 0000                    READING WORKBOOK/WORKSHEET:  ALMOST EVERY DAY
002 WB/WS-T2 (2          ) 1000                    READING WORKBOOK/WORKSHEET:  ONCE OR TWICE A WEEK
003 WB/WS-T3 (3          ) 0100                    READING WORKBOOK/WORKSHEET:  ONCE OR TWICE A MONTH
004 WB/WS-T4 (4          ) 0010                    READING WORKBOOK/WORKSHEET:  NEVER OR HARDLY EVER
005 WB/WS-T? (M          ) 0001                    READING WORKBOOK/WORKSHEET:  MISSING, MISSING NOT
                                                   APPLY

CONDITIONING VARIABLE ID:  TSUB0018
DESCRIPTION:               HOW OFTEN DO YOU ASK STUDENTS TO READ SILENTLY?
GRADES/ASSESSMENTS:        N04, S04, N08
CONDITIONING VAR LABEL:    T_SILENT
NAEP ID:                   T046706                 TOTAL NUMBER OF SPECIFIED CONTRASTS:    5
TYPE OF CONTRAST:          CLASS                   NUMBER OF INDEPENDENT CONTRASTS:        4

001 SILNT-T1 (1          ) 0000                    READ SILENTLY: ALMOST EVERY DAY
002 SILNT-T2 (2          ) 1000                    READ SILENTLY: ONCE OR TWICE A WEEK
003 SILNT-T3 (3          ) 0100                    READ SILENTLY: ONCE OR TWICE A MONTH
004 SILNT-T4 (4          ) 0010                    READ SILENTLY: NEVER OR HARDLY EVER
005 SILNT-T? (M          ) 0001                    READ SILENTLY: MISSING, DOES NOT APPLY

CONDITIONING VARIABLE ID:  TSUB0019
DESCRIPTION:               HOW OFTEN DO YOU GIVE STUDENTS TIME TO READ BOOKS OF THEIR OWN CHOOSING?
GRADES/ASSESSMENTS:        N04, S04, N08
CONDITIONING VAR LABEL:    T_OWNBKS
NAEP ID:                   T046707                 TOTAL NUMBER OF SPECIFIED CONTRASTS:    5
TYPE OF CONTRAST:          CLASS                   NUMBER OF INDEPENDENT CONTRASTS:        4

001 OWNBK-T1 (1          ) 0000                    BOOKS CHOSEN YOURSELF:  ALMOST EVERY DAY
002 OWNBK-T2 (2          ) 1000                    BOOKS CHOSEN YOURSELF:  OR TWICE A WEEK
003 OWNBK-T3 (3          ) 0100                    BOOKS CHOSEN YOURSELF:  ONCE OR TWICE A MONTH
004 OWNBK-T4 (4          ) 0010                    BOOKS CHOSEN YOURSELF:  NEVER OR HARDLY EVER
005 OWNBK-T? (M          ) 0001                    BOOKS CHOSEN YOURSELF:  MISSING, DOES NOT APPLY

CONDITIONING VARIABLE ID:  TSUB0020
DESCRIPTION:               HOW OFTEN DO YOU ASK STUDENTS TO DO A GROUP ACTIVITY/PROJECT ABOUT WHAT THEY HAVE
                           READ?
GRADES/ASSESSMENTS:        N04, S04, N08
CONDITIONING VAR LABEL:    T_PROJCT
NAEP ID:                   T046709                 TOTAL NUMBER OF SPECIFIED CONTRASTS:    5
TYPE OF CONTRAST:          CLASS                   NUMBER OF INDEPENDENT CONTRASTS:        4

001 PRJCT-T1 (1          ) 0000                    PROJECT ABOUT READING:  ALMOST EVERY DAY
002 PRJCT-T2 (2          ) 1000                    PROJECT ABOUT READING:  ONCE OR TWICE A WEEK
003 PRJCT-T3 (3          ) 0100                    PROJECT ABOUT READING:  ONCE OR TWICE A MONTH
```

Table C-4 (continued)

Conditioning Variables Specific to the Main Reading Assessment

```
004 PRJCT-T4 (4           ) 0010                PROJECT ABOUT READING:  NEVER OR HARDLY EVER
005 PRJCT-T? (M           ) 0001                PROJECT ABOUT READING:  MISSING, DOES NOT APPLY

CONDITIONING VARIABLE ID:  TSUB0021
DESCRIPTION:               HOW OFTEN DO YOU ASK STUDENTS TO DISCUSS DIFFERENT INTERPRETATIONS OF WHAT THEY
                           HAVE READ?
GRADES/ASSESSMENTS:        N04, S04, N08
CONDITIONING VAR LABEL:    T_INTERP
NAEP ID:                   T046710              TOTAL NUMBER OF SPECIFIED CONTRASTS:     5
TYPE OF CONTRAST:          CLASS                NUMBER OF INDEPENDENT CONTRASTS:         4

001 INTRP-T1 (1           ) 0000                DISCUSS READING INTERPRETATIONS:  ALMOST EVERY DAY
002 INTRP-T2 (2           ) 1000                DISCUSS READING INTERPRETATIONS:  ONCE OR TWICE A
                                                                                  WEEK
003 INTRP-T3 (3           ) 0100                DISCUSS READING INTERPRETATIONS:  ONCE OR TWICE A
                                                                                  MONTH
004 INTRP-T4 (4           ) 0010                DISCUSS READING INTERPRETATIONS:  NEVER OR HARDLY
                                                                                  EVER
005 INTRP-T? (M           ) 0001                DISCUSS READING INTERPRETATIONS:  MISSING, DOES
                                                                                  NOT APPLY

CONDITIONING VARIABLE ID:  TSUB0022
DESCRIPTION:               HOW OFTEN DO YOU ASK STUDENTS TO EXPLAIN OR SUPPORT THEIR UNDERSTANDING OF WHAT
                           THEY HAVE READ?
GRADES/ASSESSMENTS:        N04, S04, N08
CONDITIONING VAR LABEL:    T_EXPLAN
NAEP ID:                   T046711              TOTAL NUMBER OF SPECIFIED CONTRASTS:     5
TYPE OF CONTRAST:          CLASS                NUMBER OF INDEPENDENT CONTRASTS:         4

001 EXPLA-T1 (1           ) 0000                EXPLAIN/SUPPORT READING:  ALMOST EVERY DAY
002 EXPLA-T2 (2           ) 1000                EXPLAIN/SUPPORT READING:  ONCE OR TWICE A WEEK
003 EXPLA-T3 (3           ) 0100                EXPLAIN/SUPPORT READING:  ONCE OR TWICE A MONTH
004 EXPLA-T4 (4           ) 0010                EXPLAIN/SUPPORT READING:  NEVER OR HARDLY EVER
005 EXPLA-T? (M           ) 0001                EXPLAIN/SUPPORT READING:  MISSING, DOES NOT APPLY

CONDITIONING VARIABLE ID:  TSUB0023
DESCRIPTION:               HOW OFTEN DO YOU GIVE READING QUIZZES OR TESTS?
GRADES/ASSESSMENTS:        N04, S04, N08
CONDITIONING VAR LABEL:    T_QUIZES
NAEP ID:                   T046712              TOTAL NUMBER OF SPECIFIED CONTRASTS:     5
TYPE OF CONTRAST:          CLASS                NUMBER OF INDEPENDENT CONTRASTS:         4

001 QUIZS-T1 (1           ) 0000                READING QUIZZES OR TESTS:  ALMOST EVERY DAY
002 QUIZS-T2 (2           ) 1000                READING QUIZZES OR TESTS:  ONCE OR TWICE A WEEK
003 QUIZS-T3 (3           ) 0100                READING QUIZZES OR TESTS:  ONCE OR TWICE A MONTH
004 QUIZS-T4 (4           ) 0010                READING QUIZZES OR TESTS:  NEVER OR HARDLY EVER
005 QUIZA-T? (M           ) 0001                READING QUIZZES OR TESTS:  MISSING, DOES NOT APPLY

CONDITIONING VARIABLE ID:  TSUB0024
DESCRIPTION:               HOW OFTEN DO YOU USE MOVIES, VIDEOS, FILMSTRIPS, TV, TAPES, CDS, OR RECORDS?
GRADES/ASSESSMENTS:        N04, S04, N08
CONDITIONING VAR LABEL:    T_MOVIES
NAEP ID:                   T046713              TOTAL NUMBER OF SPECIFIED CONTRASTS:     5
TYPE OF CONTRAST:          CLASS                NUMBER OF INDEPENDENT CONTRASTS:         4

001 MOVIE-T1 (1           ) 0000                MOVIES, VIDEOS, TV, CDS:  ALMOST EVERY DAY
002 MOVIE-T2 (2           ) 1000                MOVIES, VIDEOS, TV, CDS:  ONCE OR TWICE A WEEK
003 MOVIE-T3 (3           ) 0100                MOVIES, VIDEOS, TV, CDS:  ONCE OR TWICE A MONTH
004 MOVIE-T4 (4           ) 0010                MOVIES, VIDEOS, TV, CDS:  NEVER OR HARDLY EVER
005 MOVIE-T? (M           ) 0001                MOVIES, VIDEOS, TV, CDS:  MISSING, DOES NOT APPLY

CONDITIONING VARIABLE ID:  TSUB0025
```

BEST COPY AVAILABLE

Table C-4 (continued)

Conditioning Variables Specific to the Main Reading Assessment

```
DESCRIPTION:                 HOW OFTEN DO YOU HAVE STUDENTS WORK ON COMPUTERS FOR READING INSTRUCTION?
GRADES/ASSESSMENTS:          N08
CONDITIONING VAR LABEL:      T_COMPTR
NAEP ID:                     T046714                   TOTAL NUMBER OF SPECIFIED CONTRASTS:     5
TYPE OF CONTRAST:            CLASS                     NUMBER OF INDEPENDENT CONTRASTS:         4

001 CMPTR-T1 (1          ) 0000                        COMPUTERS FOR READING:  ALMOST EVERY DAY
002 CMPTR-T2 (2          ) 1000                        COMPUTERS FOR READING:  ONCE OR TWICE A WEEK
003 CMPTR-T3 (3          ) 0100                        COMPUTERS FOR READING:  ONCE OR TWICE A MONTH
004 CMPTR-T4 (4          ) 0010                        COMPUTERS FOR READING:  NFVER OR HARDLY EVER
005 CMPTR-T? (M          ) 0001                        COMPUTERS FOR READING:  MISSING, DOES NOT APPLY

CONDITIONING VARIABLE ID:    TSUB0026
DESCRIPTION:                 HOW OFTEN DO YOU USE MULTIPLE-CHOICE TESTS TO ASSESS STUDENTS IN READING?
GRADES/ASSESSMENTS:          N04, S04, N08
CONDITIONING VAR LABEL:      MC TESTS
NAEP ID:                     T047001                   TOTAL NUMBER OF SPECIFIED CONTRASTS:     5
TYPE OF CONTRAST:            CLASS                     NUMBER OF INDEPENDENT CONTRASTS:         4

001 MCTEST-1 (1          ) 0000                        MULTIPLE-CHOICE TESTS:  ONCE OR TWICE A WEEK
002 MCTEST-2 (2          ) 1000                        MULTIPLE-CHOICE TESTS:  ONCE OR TWICE A MONTH
003 MCTEST-3 (3          ) 0100                        MULTIPLE-CHOICE TESTS:  ONCE OR TWICE A YEAR
004 MCTEST-4 (4          ) 0010                        MULTIPLE-CHOICE TESTS:  NEVER OR HARDLY EVER
005 MCTEST-? (M          ) 0001                        MULTIPLE-CHOICE TESTS:  MISSING, DOES NOT APPLY

CONDITIONING VARIABLE ID:    TSUB0027
DESCRIPTION:                 HOW OFTEN DO YOU USE SHORT-ANSWER TESTS TO ASSESS STUDENTS IN READING?
GRADES/ASSESSMENTS:          N04, S04, N08
CONDITIONING VAR LABEL:      SA TESTS
NAEP ID:                     T047002                   TOTAL NUMBER OF SPECIFIED CONTRASTS:     5
TYPE OF CONTRAST:            CLASS                     NUMBER OF INDEPENDENT CONTRASTS:         4

001 SATEST-1 (1          ) 0000                        SHORT-ANSWER TESTS:  ONCE OR TWICE A WEEK
002 SATEST-2 (2          ) 1000                        SHORT-ANSWER TESTS:  ONCE OR TWICE A MONTH
003 SATEST-3 (3          ) 0100                        SHORT-ANSWER TESTS:  ONCE OR TWICE A YEAR
004 SATEST-4 (4          ) 0010                        SHORT-ANSWER TESTS:  NEVER OR HARDLY EVER
005 SATEST-? (M          ) 0001                        SHORT-ANSWER TESTS:  MISSING, DOES NOT APPLY

CONDITIONING VARIABLE ID:    TSUB0028
DESCRIPTION:                 HOW OFTEN DO YOU USE WRITING PARAGRAPHS TO ASSESS STUDENTS IN READING?
GRADES/ASSESSMENTS:          N04, S04, N08
CONDITIONING VAR LABEL:      WRI TEST
NAEP ID:                     T047003                   TOTAL NUMBER OF SPECIFIED CONTRASTS:     5
TYPE OF CONTRAST:            CLASS                     NUMBER OF INDEPENDENT CONTRASTS:         4

001 WRITST-1 (1          ) 0000                        ASSESS BY WRITING PARAGRAPHS:  ONCE OR TWICE A
                                                                                      WEEK
002 WRITST-2 (2          ) 1000                        ASSESS BY WRITING PARAGRAPHS:  ONCE OR TWICE A
                                                                                      MONTH
003 WRITST-3 (3          ) 0100                        ASSESS BY WRITING PARAGRAPHS:  ONCE OR TWICE A
                                                                                      YEAR
004 WRITST-4 (4          ) 0010                        ASSESS BY WRITING PARAGRAPHS:  NEVER OR HARDLY
                                                                                      EVER
005 WRITST-? (M          ) 0001                        ASSESS BY WRITING PARAGRAPHS:  MISSING, DOES NOT
                                                                                      APPLY

CONDITIONING VARIABLE II     TSUB0029
DESCRIPTION:                  OW OFTEN DO YOU USE INDIVIDUAL OR GROUP PROJECTS/PRESENTATIONS TO ASSESS STUDENTS
                             IN READING?
GRADES/ASSESSMENTS:          N04, S04, N08
CONDITIONING VAR LABEL:      PRJ TEST
NAEP ID:                     T047006                   TOTAL NUMBER OF SPECIFIED CONTRASTS:     5
```

Table C-4 (continued)

Conditioning Variables Specific to the Main Reading Assessment

```
TYPE OF CONTRAST:          CLASS                          NUMBER OF INDEPENDENT CONTRASTS:        4

001 PRJTST-1 (1          ) 0000                           ASSESS BY PROJECTS/PRESENTATIONS:  ONCE OR TWICE A
                                                                                             WEEK
002 PRJTST-2 (2          ) 1000                           ASSESS BY PROJECTS/PRESENTATIONS:  ONCE OR TWICE A
                                                                                             MONTH
003 PRJTST-3 (3          ) 0100                           ASSESS BY PROJECTS/PRESENTATIONS:  ONCE OR TWICE A
                                                                                             YEAR
004 PRJTST-4 (4          ) 0010                           ASSESS BY PROJECTS/PRESENTATIONS:  NEVER OR HARDLY
                                                                                             EVER
005 PRJTST-? (M          ) 0001                           ASSESS BY PROJECTS/PRESENTATIONS:  MISSING, DOES
                                                                                             NOT APPLY

CONDITIONING VARIABLE ID:  TSUB0030
DESCRIPTION:               HOW OFTEN DO YOU USE READING PORTFOLIOS TO ASSESS STUDENTS IN READING?
GRADES/ASSESSMENTS:        N04, S04, N08
CONDITIONING VAR LABEL:    PRT TEST
NAEP ID:                   T047007                        TOTAL NUMBER OF SPECIFIED CONTRASTS:    5
TYPE OF CONTRAST:          CLASS                          NUMBER OF INDEPENDENT CONTRASTS:        4

001 PRTTST-1 (1          ) 0000                           ASSESS BY READING PORTFOLIOS:  ONCE OR TWICE A
                                                                                         WEEK
002 PRTTST-2 (2          ) 1000                           ASSESS BY READING PORTFOLIOS:  ONCE OR TWICE A
                                                                                         MONTH
003 PRTTST-3 (3          ) 0100                           ASSESS BY READING PORTFOLIOS:  ONCE OR TWICE A
                                                                                         YEAR
004 PRTTST-4 (4          ) 0010                           ASSESS BY READING PORTFOLIOS:  NEVER OR HARDLY
                                                                                         EVER
005 PRTTST-? (M          ) 0001                           ASSESS BY READING PORTFOLIOS:  MISSING, DOES NOT
                                                                                         APPLY

CONDITIONING VARIABLE ID:  TSUB0031
DESCRIPTION:               HOW OFTEN DO YOU SEND OR TAKE THE CLASS TO THE LIBRARY?
GRADES/ASSESSMENTS:        N04, S04, N08
CONDITIONING VAR LABEL:    CLAS2LIB
NAEP ID:                   T047101                        TOTAL NUMBER OF SPECIFIED CONTRASTS:    6
TYPE OF CONTRAST:          CLASS                          NUMBER OF INDEPENDENT CONTRASTS:        5

001 CLSLIB-1 (1          ) 00000                          TAKE CLASS TO LIBRARY:  ALMOST EVERY DAY
002 CLSLIB-2 (2          ) 10000                          TAKE CLASS TO LIBRARY:  ONCE OR TWICE A WEEK
003 CLSLIB-3 (3          ) 01000                          TAKE CLASS TO LIBRARY:  ONCE OR TWICE A MONTH
004 CLSLIB-4 (4          ) 00100                          TAKE CLASS TO LIBRARY:  NEVER OR HARDLY EVER
005 CLSLIB-5 (5          ) 00010                          TAKE CLASS TO LIBRARY:  THERE IS NO LIBRARY
006 CLSLIB-? (M          ) 00001                          TAKE CLASS TO LIBRARY:  MISSING, DOES NOT APPLY

CONDITIONING VARIABLE ID:  TSUB0032
DESCRIPTION:               HOW OFTEN DO YOU ASSIGN STUDENTS TO READ A BOOK FROM THE LIBRARY?
GRADES/ASSESSMENTS:        N04, S04, N08
CONDITIONING VAR LABEL:    LIB BOOK
NAEP ID:                   T047102                        TOTAL NUMBER OF SPECIFIED CONTRASTS:    6
TYPE OF CONTRAST:          CLASS                          NUMBER OF INDEPENDENT CONTRASTS:        5

001 CLALIB-1 (1          ) 00000                          ASSIGN STUDENTS TO READ LIBRARY BOOK:  ALMOST
                                                          EVERY DAY
002 CLALIB-2 (2          ) 10000                          ASSIGN STUDENTS TO READ LIBRARY BOOK:  ONCE OR
                                                          TWICE A WEEK
003 CLALIB-3 (3          ) 01000                          ASSIGN STUDENTS TO READ LIBRARY BOOK:  ONCE OR
                                                          TWICE A MONTH
004 CLALIB-4 (4          ) 00100                          ASSIGN STUDENTS TO READ LIBRARY BOOK:  NEVER OR
                                                          HARDLY EVER
005 CLALIB-5 (5          ) 00010                          ASSIGN STUDENTS TO READ LIBRARY BOOK:  THERE IS NO
                                                          LIBRARY
```

Table C-4 (continued)

Conditioning Variables Specific to the Main Reading Assessment

```
006 CLALIB-? (M          ) 00001                 ASSIGN STUDENTS TO READ LIBRARY BOOK:  MISSING,
                                                 DOES NOT APPLY

CONDITIONING VARIABLE ID:  TSUB0033
DESCRIPTION:               ARE COMPUTERS AVAILABLE FOR USE BY STUDENTS IN READING CLASS?
GRADES/ASSESSMENTS:        N04, S04, N08
CONDITIONING VAR LABEL:    COMP4RED
NAEP ID:                   T047201               TOTAL NUMBER OF SPECIFIED CONTRASTS:     4
TYPE OF CONTRAST:          CLASS                 NUMBER OF INDEPENDENT CONTRASTS:         3

001 COMP-NA  (1          ) 000                   COMPUTERS IN READING CLASS:  NOT AVAILABLE
002 COMP-DIF (2          ) 100                   COMPUTERS IN READING CLASS:  AVAILABLE BUT
                                                 DIFFICULT TO ACCESS
003 COMP-AVL (3          ) 010                   COMPUTERS IN READING CLASS:  AVAILABLE IN THE
                                                 CLASSROOM
004 COMP-?   (M          ) 001                   COMPUTERS IN READING CLASS:  MISSING, DOES NOT
                                                 APPLY
```

Table C-5

Conditioning Variables Specific to the U.S. History Assessment

```
CONDITIONING VARIABLE ID:   SUBJ0001
DESCRIPTION:                HOW OFTEN DO YOU USUALLY HAVE SOCIAL STUDIES OR HISTORY CLASSES IN SCHOOL?

GRADES/ASSESSMENTS:         N04
CONDITIONING VAR LABEL:     OFTEN
NAEP ID:                    H802901                          TOTAL NUMBER OF SPECIFIED CONTRASTS:     6
TYPE OF CONTRAST:           CLASS                            NUMBER OF INDEPENDENT CONTRASTS:         5


001 OFT_DAY  (1           ) 00000                            HOW OFTEN HISTORY:  EVERY DAY
002 OFT_3-4W (2           ) 10000                            HOW OFTEN HISTORY:  3-4 TIMES A WEEK
003 OFT_1-2W (3           ) 01000                            HOW OFTEN HISTORY:  1-2 TIMES A WEEK
004 OFT_<1W  (4           ) 00100                            HOW OFTEN HISTORY:  LESS THAN ONCE A WEEK
005 OFT_NEV  (5           ) 00010                            HOW OFTEN HISTORY:  NEVER, OR, HARDLY EVER
006 OFT_MIS  (M           ) 00001                            HOW OFTEN HISTORY:  MISSING

CONDITIONING VARIABLE ID:   SUBJ0002
DESCRIPTION:                HAVE YOU EVER STUDIED THE HISTORY OF THE UNITED STATES IN SCHOOL?
GRADES/ASSESSMENTS:         N04
CONDITIONING VAR LABEL:     HISTORY?
NAEP ID:                    H803001                          TOTAL NUMBER OF SPECIFIED CONTRASTS:     3
TYPE OF CONTRAST:           CLASS                            NUMBER OF INDEPENDENT CONTRASTS:         2

001 YES      (1           ) 00STUDIED HISTORY?: YES
002 NO       (2           ) 10STUDIED HISTORY?: NO
003 MISSING  (M           ) 01STUDIED HISTORY?: MISSING

CONDITIONING VARIABLE ID:   SUBJ0003
DESCRIPTION:                DO EITHER YOU OR YOUR TEACHER SAVE YOUR HISTORY OR SOCIAL STUDIES WORK IN A FOLDER
                            OR A PORTFOLIO?
GRADES/ASSESSMENTS:         N04, N08, N12
CONDITIONING VAR LABEL:     PORTFOLI
NAEP ID:                    H803101                          TOTAL NUMBER OF SPECIFIED CONTRASTS:     3
TYPE OF CONTRAST:           CLASS                            NUMBER OF INDEPENDENT CONTRASTS:         2

001 PORT-YES (1           ) 00PORTFOLIO: YES
002 PORT--NO (2           ) 10PORTFOLIO: NO
003 PORT-MIS (M           ) 01PORTFOLIO: MISSING

CONDITIONING VARIABLE ID:   SUBJ0004
DESCRIPTION:                HOW OFTEN DO YOU READ MATERIAL FROM A TEXTBOOK?
GRADES/ASSESSMENTS:         N04, N08, N12
CONDITIONING VAR LABEL:     RDTEXTBK
NAEP ID:                    H803201                          TOTAL NUMBER OF SPECIFIED CONTRASTS:     6
TYPE OF CONTRAST:           CLASS                            NUMBER OF INDEPENDENT CONTRASTS:         5

001 RDTEXT-1 (1           ) 00000                            HOW OFTEN TEXTBOOK: ABOUT EVERY DAY
002 RDTEXT-2 (2           ) 10000                            HOW OFTEN TEXTBOOK: ONCE OR TWICE A WEEK
003 RDTEXT-3 (3           ) 01000                            HOW OFTEN TEXTBOOK: ONCE OR TWICE A MONTH
004 RDTEXT-4 (4           ) 00100                            HOW OFTEN TEXTBOOK: A FEW TIMES A YEAR
005 RDTEXT-5 (5           ) 00010                            HOW OFTEN TEXTBOOK: NEVER
006 RDTEXT-? (M           ) 00001                            HOW OFTEN TEXTBOOK: MISSING

CONDITIONING VARIABLE ID:   SUBJ0005
DESCRIPTION:                HOW OFTEN DO YOU READ EXTRA MATERIAL NOT IN A TEXTBOOK?
GRADES/ASSESSMENTS:         N04, N08, N12
CONDITIONING VAR LABEL:     RDEXTMAT
NAEP ID:                    H803202                          TOTAL NUMBER OF SPECIFIED CONTRASTS:     6
TYPE OF CONTRAST:           CLASS                            NUMBER OF INDEPENDENT CONTRASTS:         5

001 EXTMAT-1 (1           ) 00000                            HOW OFTEN EXTRA MATERIAL: ABOUT EVERY DAY
002 EXTMAT-2 (2           ) 10000                            HOW OFTEN EXTRA MATERIAL: ONCE OR TWICE A WEEK
003 EXTMAT-3 (3           ) 01000                            HOW OFTEN EXTRA MATERIAL: ONCE OR TWICE A MONTH
```

Table C-5 (continued)

Conditioning Variables Specific to the U.S. History Assessment

```
004 EXTMAT-4 (4          ) 00100                    HOW OFTEN EXTRA MATERIAL: A FEW TIMES A YEAR
005 EXTMAT-5 (5          ) 00010                    HOW OFTEN EXTRA MATERIAL: NEVER
006 EXTMAT-? (M          ) 00001                    HOW OFTEN EXTRA MATERIAL: MISSING

CONDITIONING VARIABLE ID:  SUBJ0006
DESCRIPTION:               HOW OFTEN DO YOU READ WRITINGS OF HISTORIC PEOPLE?
GRADES/ASSESSMENTS:        N04, N08, N12
CONDITIONING VAR LABEL:    RDHISPEO
NAEP ID:                   H803203                  TOTAL NUMBER OF SPECIFIED CONTRASTS:    6
TYPE OF CONTRAST:          CLASS                    NUMBER OF INDEPENDENT CONTRASTS:        5

001 HISPEO-1 (1          ) 00000                    HOW OFTEN WRITINGS OF HISTORIC PEOPLE: ABOUT EVERY
                                                    DAY
002 HISPEO-2 (2          ) 10000                    HOW OFTEN WRITINGS OF HISTORIC PEOPLE: ONCE OR
                                                    TWICE A WEEK
003 HISPEO-3 (3          ) 01000                    HOW OFTEN WRITINGS OF HISTORIC PEOPLE: ONCE OR
                                                    TWICE A MONTH
004 HISPEO-4 (4          ) 00100                    HOW OFTEN WRITINGS OF HISTORIC PEOPLE: A FEW TIMES
                                                    A YEAR
005 HISPEO-5 (5          ) 00010                    HOW OFTEN WRITINGS OF HISTORIC PEOPLE: NEVER
006 HISPEO-? (M          ) 00001                    HOW OFTEN WRITINGS OF HISTORIC PEOPLE: MISSING

CONDITIONING VARIABLE ID:  SUBJ0007
DESCRIPTION:               HOW OFTEN DO YOU DISCUSS MATERIAL STUDIED?
GRADES/ASSESSMENTS:        N04, N08, N12
CONDITIONING VAR LABEL:    DISCMAT
NAEP ID:                   H803204                  TOTAL NUMBER OF SPECIFIED CONTRASTS:    6
TYPE OF CONTRAST:          CLASS                    NUMBER OF INDEPENDENT CONTRASTS:        5

001 DISMAT-1 (1          ) 00000                    HOW OFTEN DISCUSS MATERIAL STUDIED: ABOUT EVERY
                                                    DAY
002 DISMAT-2 (2          ) 10000                    HOW OFTEN DISCUSS MATERIAL STUDIED: ONCE OR TWICE
                                                    A WEEK
003 DISMAT-3 (3          ) 01000                    HOW OFTEN DISCUSS MATERIAL STUDIED: ONCE OR TWICE
                                                    A MONTH
004 DISMAT-4 (4          ) 00100                    HOW OFTEN DISCUSS MATERIAL STUDIED: A FEW TIMES A
                                                    YEAR
005 DISMAT-5 (5          ) 00010                    HOW OFTEN DISCUSS MATERIAL STUDIED: NEVER

006 DISMAT-? (M          ) 00001                    HOW OFTEN DISCUSS MATERIAL STUDIED: MISSING

CONDITIONING VARIABLE ID:  SUBJ0008
DESCRIPTION:               HOW OFTEN DO YOU WRITE SHORT ANSWERS TO QUESTIONS?
GRADES/ASSESSMENTS:        N04, N08, N12
CONDITIONING VAR LABEL:    WSHANS
NAEP ID:                   H803205                  TOTAL NUMBER OF SPECIFIED CONTRASTS:    6
TYPE OF CONTRAST:          CLASS                    NUMBER OF INDEPENDENT CONTRASTS:        5

001 WSHANS-1 (1          ) 00000                    HOW OFTEN WRITE SHORT ANSWERS: ABOUT EVERY DAY

002 WSHANS-2 (2          ) 10000                    HOW OFTEN WRITE SHORT ANSWERS: ONCE OR TWICE A
                                                    WEEK
003 WSHANS-3 (3          ) 01000                    HOW OFTEN WRITE SHORT ANSWERS: ONCE OR TWICE A
                                                    MONTH
004 WSHANS-4 (4          ) 00100                    HOW OFTEN WRITE SHORT ANSWERS: A FEW TIMES A YEAR
005 WSHANS-5 (5          ) 00010                    HOW OFTEN WRITE SHORT ANSWERS: NEVER
006 WSHANS-? (M          ) 00001                    HOW OFTEN WRITE SHORT ANSWERS: MISSING

CONDITIONING VARIABLE ID:  SUBJ0009
DESCRIPTION:               HOW OFTEN DO YOU WRITE A REPORT?
GRADES/ASSESSMENTS:        N04, N08, N12
CONDITIONING VAR LABEL:    WREPRT
NAEP ID:                   H803206                  TOTAL NUMBER OF SPECIFIED CONTRASTS:    6
```

Table C-5 (continued)

Conditioning Variables Specific to the U.S. History Assessment

```
TYPE OF CONTRAST:          CLASS                          NUMBER OF INDEPENDENT CONTRASTS:        5

001 WREPRT-1 (1           ) 00000                          HOW OFTEN WRITE A REPORT: ABOUT EVERY DAY
002 WREPRT-2 (2           ) 10000                          HOW OFTEN WRITE A REPORT: ONCE OR TWICE A WEEK
003 WREPRT-3 (3           ) 01000                          HOW OFTEN WRITE A REPORT: ONCE OR TWICE A MONTH
004 WREPRT-4 (4           ) 00100                          HOW OFTEN WRITE A REPORT: A FEW TIMES A YEAR
005 WREPRT-5 (5           ) 00010                          HOW OFTEN WRITE A REPORT: NEVER
006 WREPRT-? (M           ) 00001                          HOW OFTEN WRITE A REPORT: MISSING

CONDITIONING VARIABLE ID:  SUBJ0010
DESCRIPTION:               HOW OFTEN DO YOU WORK ON A GROUP PROJECT?
GRADES/ASSESSMENTS:        N04, N08, N12
CONDITIONING VAR LABEL:    GRPPROJ
NAEP ID:                   H803207                        TOTAL NUMBER OF SPECIFIED CONTRASTS:    6
TYPE OF CONTRAST:          CLASS                          NUMBER OF INDEPENDENT CONTRASTS:        5

001 GRPPRJ-1 (1           ) 00000                          HOW OFTEN WORK ON A GROUP PROJECT: ABOUT EVERY DAY
002 GRPPRJ-2 (2           ) 10000                          HOW OFTEN WORK ON A GROUP PROJECT: ONCE OR TWICE A
                                                           WEEK
003 GRPPRJ-3 (3           ) 01000                          HOW OFTEN WORK ON A GROUP PROJECT: ONCE OR TWICE A
                                                           MONTH
004 GRPPRJ-4 (4           ) 00100                          HOW OFTEN WORK ON A GROUP PROJECT: A FEW TIMES A
                                                           YEAR
005 GRPPRJ-5 (5           ) 00010                          HOW OFTEN WORK ON A GROUP PROJECT: NEVER
006 GRPPRJ-? (M           ) 00001                          HOW OFTEN WORK ON A GROUP PROJECT: MISSING

CONDITIONING VARIABLE ID:  SUBJ0011
DESCRIPTION:               HOW OFTEN DO YOU GIVE A REPORT ON THE TOPIC BEING STUDIED?
GRADES/ASSESSMENTS:        N04, N08, N12
CONDITIONING VAR LABEL:    GIVREP
NAEP ID:                   H803208                        TOTAL NUMBER OF SPECIFIED CONTRASTS:    6
TYPE OF CONTRAST:          CLASS                          NUMBER OF INDEPENDENT CONTRASTS:        5

001 GIVREP-1 (1           ) 00000                          HOW OFTEN REPORT ON TOPIC: ABOUT EVERY DAY
002 GIVREP-2 (2           ) 10000                          HOW OFTEN REPORT ON TOPIC: ONCE OR TWICE A WEEK
003 GIVREP-3 (3           ) 01000                          HOW OFTEN REPORT ON TOPIC: ONCE OR TWICE A MONTH
004 GIVREP-4 (4           ) 00100                          HOW OFTEN REPORT ON TOPIC: A FEW TIMES A YEAR
005 GIVREP-5 (5           ) 00010                          HOW OFTEN REPORT ON TOPIC: NEVER
006 GIVREP-? (M           ) 00001                          HOW OFTEN REPORT ON TOPIC: MISSING

CONDITIONING VARIABLE ID:  SUBJ0012
DESCRIPTION:               HOW OFTEN DO YOU USE MAPS AND GLOBES?
GRADES/ASSESSMENTS:        N04, N08, N12
CONDITIONING VAR LABEL:    MAPSGL
NAEP ID:                   H803209                        TOTAL NUMBER OF SPECIFIED CONTRASTS:    6
TYPE OF CONTRAST:          CLASS                          NUMBER OF INDEPENDENT CONTRASTS:        5

001 MAPSGL-1 (1           ) 00000                          HOW OFTEN USE MAPS AND GLOBES: ABOUT EVERY DAY
002 MAPSGL-2 (2           ) 10000                          HOW OFTEN USE MAPS AND GLOBES: ONCE OR TWICE A
                                                           WEEK
003 MAPSGL-3 (3           ) 01000                          HOW OFTEN USE MAPS AND GLOBES: ONCE OR TWICE A
                                                           MONTH
004 MAPSGL-4 (4           ) 00100                          HOW OFTEN USE MAPS AND GLOBES: A FEW TIMES A YEAR
005 MAPSGL-5 (5           ) 00010                          HOW OFTEN USE MAPS AND GLOBES: NEVER
006 MAPSGL-? (M           ) 00001                          HOW OFTEN USE MAPS AND GLOBES: MISSING

CONDITIONING VARIABLE ID:  SUBJ0013
DESCRIPTION:               HOW OFTEN DO YOU WATCH MOVIES, VIDEOS, OR FILMSTRIPS?
GRADES/ASSESSMENTS:        N04, N08, N12
CONDITIONING VAR LABEL:    VIDEO
NAEP ID:                   H803210                        TOTAL NUMBER OF SPECIFIED CONTRASTS:    6
TYPE OF CONTRAST:          CLASS                          NUMBER OF INDEPENDENT CONTRASTS:        5
```

BEST COPY AVAILABLE

Table C-5 (continued)

Conditioning Variables Specific to the U.S. History Assessment

```
001 VIDEO--1 (1          ) 00000                   HOW OFTEN WATCH VIDEOS: ABOUT EVERY DAY
002 VIDEO--2 (2          ) 10000                   HOW OFTEN WATCH VIDEOS: ONCE OR TWICE A WEEK
003 VIDEO--3 (3          ) 01000                   HOW OFTEN WATCH VIDEOS: ONCE OR TWICE A MONTH
004 VIDEO--4 (4          ) 00100                   HOW OFTEN WATCH VIDEOS: A FEW TIMES A YEAR
005 VIDEO--5 (5          ) 00010                   HOW OFTEN WATCH VIDEOS: NEVER
006 VIDEO--? (M          ) 00001                   HOW OFTEN WATCH VIDEOS: MISSING

CONDITIONING VARIABLE ID:  SUBJ0014
DESCRIPTION:               HOW OFTEN DO YOU USE A COMPUTER?
GRADES/ASSESSMENTS:        N04, N08, N12
CONDITIONING VAR LABEL:    COMPUTER
NAEP ID:                   H803211                 TOTAL NUMBER OF SPECIFIED CONTRASTS:     6
TYPE OF CONTRAST:          CLASS                   NUMBER OF INDEPENDENT CONTRASTS:         5

001 COMPUT-1 (1          ) 00000                   HOW OFTEN USE A COMPUTER: ABOUT EVERY DAY
002 COMPUT-2 (2          ) 10000                   HOW OFTEN USE A COMPUTER: ONCE OR TWICE A WEEK
003 COMPUT-3 (3          ) 01000                   HOW OFTEN USE A COMPUTER: ONCE OR TWICE A MONTH
004 COMPUT-4 (4          ) 00100                   HOW OFTEN USE A COMPUTER: A FEW TIMES A YEAR
005 COMPUT-5 (5          ) 00010                   HOW OFTEN USE A COMPUTER: NEVER
006 COMPUT-? (M          ) 00001                   HOW OFTEN USE A COMPUTER: MISSING

CONDITIONING VARIABLE ID:  SUBJ0015
DESCRIPTION:               HOW OFTEN DO YOU TAKE A TEST OR A QUIZ?
GRADES/ASSESSMENTS:        N04, N08, N12
CONDITIONING VAR LABEL:    QUIZ
NAEP ID:                   H803212                 TOTAL NUMBER OF SPECIFIED CONTRASTS:     6
TYPE OF CONTRAST:          CLASS                   NUMBER OF INDEPENDENT CONTRASTS:         5

001 QUIZ---1 (1          ) 00000                   HOW OFTEN TAKE A TEST OR A QUIZ: ABOUT EVERY DAY
002 QUIZ---2 (2          ) 10000                   HOW OFTEN TAKE A TEST OR A QUIZ: ONCE OR TWICE A
                                                   WEEK
003 QUIZ---3 (3          ) 01000                   HOW OFTEN TAKE A TEST OR A QUIZ: ONCE OR TWICE A
                                                   MONTH
004 QUIZ---4 (4          ) 00100                   HOW OFTEN TAKE A TEST OR A QUIZ: A FEW TIMES A
                                                   YEAR
005 QUIZ---5 (5          ) 00010                   HOW OFTEN TAKE A TEST OR A QUIZ: NEVER
006 QUIZ---? (M          ) 00001                   HOW OFTEN TAKE A TEST OR A QUIZ: MISSING

CONDITIONING VARIABLE ID:  SUBJ0016
DESCRIPTION:               HOW OFTEN GO ON FIELD TRIPS OR HAVE OUTSIDE SPEAKERS?
GRADES/ASSESSMENTS:        N04, N08, N12
CONDITIONING VAR LABEL:    TRIPS
NAEP ID:                   H803213                 TOTAL NUMBER OF SPECIFIED CONTRASTS:     6
TYPE OF CONTRAST:          CLASS                   NUMBER OF INDEPENDENT CONTRASTS:         5

001 TRIPS--1 (1          ) 00000                   HOW OFTEN GO ON TRIPS: ABOUT EVERY DA
002 TRIPS--2 (2          ) 10000                   HOW OFTEN GO ON TRIPS: ONCE OR TWICE A WEEK
003 TRIPS--3 (3          ) 01000                   HOW OFTEN GO ON TRIPS: ONCE OR TWICE A MONTH
004 TRIPS--4 (4          ) 00100                   HOW OFTEN GO ON TRIPS: A FEW TIMES A YEAR
005 TRIPS--5 (5          ) 00010                   HOW OFTEN GO ON TRIPS: NEVER
006 TRIPS--? (M          ) 00001                   HOW OFTEN GO ON TRIPS: MISSING

CONDITIONING VARIABLE ID:  SUBJ0017
DESCRIPTION:               HOW OFTEN WORK IN THE LIBRARY?
GRADES/ASSESSMENTS:        N04, N08, N12
CONDITIONING VAR LABEL:    LIBRARY
NAEP ID:                   H803214                 TOTAL NUMBER OF SPECIFIED CONTRASTS:     6
TYPE OF CONTRAST:          CLASS                   NUMBER OF INDEPENDENT CONTRASTS:         5

001 LIBRA--1 (1          ) 00000                   HOW OFTEN WORK IN THE LIBRARY: ABOUT EVERY DAY
002 LIBRA--2 (2          ) 10000                   HOW OFTEN WORK IN THE LIBRARY: ONCE OR TWICE A
                                                   WEEK
003 LIBRA--3 (3          ) 01000                   HOW OFTEN WORK IN THE LIBRARY: ONCE OR TWICE A
                                                   MONTH
```

Table C-5 (continued)

Conditioning Variables Specific to the U.S. History Assessment

```
004 LIBRA--4 (4         ) 00100                 HOW OFTEN WORK IN THE LIBRARY: A FEW TIMES A YEAR
005 LIBRA--5 (5         ) 00010                 HOW OFTEN WORK IN THE LIBRARY: NEVER
006 LIBRA--? (M         ) 00001                 HOW OFTEN WORK IN THE LIBRARY: MISSING

CONDITIONING VARIABLE ID:  SUBJ0018
DESCRIPTION:               DID YOU TAKE A U.S. HISTORY COURSE IN THE SIXTH GRADE?
GRADES/ASSESSMENTS:        N08
CONDITIONING VAR LABEL:    HISGR06
NAEP ID:                   H803301              TOTAL NUMBER OF SPECIFIED CONTRASTS:     4
TYPE OF CONTRAST:          CLASS                NUMBER OF INDEPENDENT CONTRASTS:         3

001 HISGR6-Y (1         ) 000                   SIXTH GRADE: YES
002 HISGR6-M (2         ) 100                   SIXTH GRADE: NO
003 HISGR6-D (3         ) 010                   SIXTH GRADE: I DON'T KNOW
004 HISGR6-? (M         ) 001                   SIXTH GRADE: MISSING

CONDITIONING VARIABLE ID:  SUBJ0019
DESCRIPTION:               DID YOU TAKE A U.S. HISTORY COURSE IN THE SEVENTH GRADE?
GRADES/ASSESSMENTS:        N08
CONDITIONING VAR LABEL:    HISGR07
NAEP ID:                   H803302              TOTAL NUMBER OF SPECIFIED CONTRASTS:     4
TYPE OF CONTRAST:          CLASS                NUMBER OF INDEPENDENT CONTRASTS:         3

001 HISGR7-Y (1         ) 000                   SEVENTH GRADE: YES
002 HISGR7-M (2         ) 100                   SEVENTH GRADE: NO
003 HISGR7-D (3         ) 010                   SEVENTH GRADE: I DON'T KNOW
004 HISGR7-? (M         ) 001                   SEVENTH GRADE: MISSING

CONDITIONING VARIABLE ID:  SUBJ0020
DESCRIPTION:               DID YOU TAKE A U.S. HISTORY COURSE IN THE NINTH GRADE?
GRADES/ASSESSMENTS:        N12
CONDITIONING VAR LABEL:    HISGR09
NAEP ID:                   H803303              TOTAL NUMBER OF SPECIFIED CONTRASTS:     4
TYPE OF CONTRAST:          CLASS                NUMBER OF INDEPENDENT CONTRASTS:         3

001 HISGR9-Y (1         ) 000                   NINTH GRADE: YES
002 HISGR9-M (2         ) 100                   NINTH GRADE: NO
003 HISGR9-D (3         ) 010                   NINTH GRADE: I DON'T KNOW
004 HISGR9-? (M         ) 001                   NINTH GRADE: MISSING

CONDITIONING VARIABLE ID:  SUBJ0021
DESCRIPTION:               DID YOU TAKE A U.S. HISTORY COURSE IN THE TENTH GRADE?
GRADES/ASSESSMENTS:        N12
CONDITIONING VAR LABEL:    HISGR10
NAEP ID:                   H803304              TOTAL NUMBER OF SPECIFIED CONTRASTS:     4
TYPE OF CONTRAST:          CLASS                NUMBER OF INDEPENDENT CONTRASTS:         3

001 HISG10-Y (1         ) 000                   TENTH GRADE: YES
002 HISG10-M (2         ) 100                   TENTH GRADE: NO
003 HISG10-D (3         ) 010                   TENTH GRADE: I DON'T KNOW
004 HISG10-? (M         ) 001                   TENTH GRADE: MISSING

CONDITIONING VARIABLE ID:  SUBJ0022
DESCRIPTION:               DID YOU TAKE A U.S. HISTORY COURSE IN THE ELEVENTH GRADE?
GRADES/ASSESSMENTS:        N12
CONDITIONING VAR LABEL:    HISGR11
NAEP ID:                   H803305              TOTAL NUMBER OF SPECIFIED CONTRASTS:     4
TYPE OF CONTRAST:          CLASS                NUMBER OF INDEPENDENT CONTRASTS:         3

001 HISG11-Y (1         ) 000                   ELEVENTH GRADE: YES
002 HISG11-M (2         ) 100                   ELEVENTH GRADE: NO
003 HISG11-D (3         ) 010                   ELEVENTH GRADE: I DON'T KNOW
004 HISG11-? (M         ) 001                   ELEVENTH GRADE: MISSING
```

Table C-5 (continued)

Conditioning Variables Specific to the U.S. History Assessment

```
CONDITIONING VARIABLE ID:  SUBJ0023
DESCRIPTION:               ARE YOU TAKING A U.S. HISTORY COURSE NOW?
GRADES/ASSESSMENTS:        N08, N12
CONDITIONING VAR LABEL:    HISTNOW
NAEP ID:                   H803401                    TOTAL NUMBER OF SPECIFIED CONTRASTS:     3
TYPE OF CONTRAST:          CLASS                      NUMBER OF INDEPENDENT CONTRASTS:         2

001 HISNOW-Y (1          ) 00HISTORY NOW: YES
002 HISNOW-N (2          ) 10HISTORY NOW: NO
003 HISNOW-? (M          ) 01HISTORY NOW: MISSING

CONDITIONING VARIABLE ID:  SUBJ0024
DESCRIPTION:               ABOUT HOW MUCH HISTORY HOMEWORK DO YOU HAVE EACH WEEK?
GRADES/ASSESSMENTS:        N08, N12
CONDITIONING VAR LABEL:    HISHWK
NAEP ID:                   H803501                    TOTAL NUMBER OF SPECIFIED CONTRASTS:     6
TYPE OF CONTRAST:          CLASS                      NUMBER OF INDEPENDENT CONTRASTS:         5

001 HISHWK-1 (1          ) 00000                      HISTORY HOMEWORK: NONE
002 HISHWK-2 (2          ) 10000                      HISTORY HOMEWORK: HALF HOUR
003 HISHWK-3 (3          ) 01000                      HISTORY HOMEWORK: ONE HOUR
004 HISHWK-4 (4          ) 00100                      HISTORY HOMEWORK: TWO HOURS
005 HISHWK-5 (5          ) 00010                      HISTORY HOMEWORK: MORE THAN 2 HOURS
006 HISHWK-? (M          ) 00001                      HISTORY HOMEWORK: MISSING

CONDITIONING VARIABLE ID:  SUBJ0025
DESCRIPTION:               HOW OFTEN HAVE YOU STUDIED THE PRE-1815 PERIOD OF U.S. HISTORY?
GRADES/ASSESSMENTS:        N08, N12
CONDITIONING VAR LABEL:    PRE1815
NAEP ID:                   H803601                    TOTAL NUMBER OF SPECIFIED CONTRASTS:     4
TYPE OF CONTRAST:          CLASS                      NUMBER OF INDEPENDENT CONTRASTS:         3

001 PR1815-1 (1          ) 000                        HOW OFTEN STUDIED PRE-1815 PERIOD: A LOT
002 PR1815-2 (2          ) 100                        HOW OFTEN STUDIED PRE-1815 PERIOD: SOME
003 PR1815-3 (3          ) 010                        HOW OFTEN STUDIED PRE-1815 PERIOD: NOT AT ALL
004 PR1815-? (M          ) 001                        HOW OFTEN STUDIED PRE-1815 PERIOD: MISSING

CONDITIONING VARIABLE ID:  SUBJ0026
DESCRIPTION:               HOW OFTEN HAVE YOU STUDIED THE 1815-1865 PERIOD OF U.S. HISTORY?
GRADES/ASSESSMENTS:        N08, N12
CONDITIONING VAR LABEL:    1815-65
NAEP ID:                   H803602                    TOTAL NUMBER OF SPECIFIED CONTRASTS:     4
TYPE OF CONTRAST:          CLASS                      NUMBER OF INDEPENDENT CONTRASTS:         3

001 181565-1 (1          ) 000                        HOW OFTEN STUDIED 1815-1865 PERIOD: A LOT
002 181565-2 (2          ) 100                        HOW OFTEN STUDIED 1815-1865 PERIOD: SOME
003 181565-3 (3          ) 010                        HOW OFTEN STUDIED 1815-1865 PERIOD: NOT AT ALL
004 181565-? (M          ) 001                        HOW OFTEN STUDIED 1815-1865 PERIOD: MISSING

CONDITIONING VARIABLE ID:  SUBJ0027
DESCRIPTION:               HOW OFTEN HAVE YOU STUDIED THE 1865-1945 PERIOD OF U.S. HISTORY?
GRADES/ASSESSMENTS:        N08, N12
CONDITIONING VAR LABEL:    1865-45
NAEP ID:                   H803603                    TOTAL NUMBER OF SPECIFIED CONTRASTS:     4
TYPE OF CONTRAST:          CLASS                      NUMBER OF INDEPENDENT CONTRASTS:         3

001 186545-1 (1          ) 000                        HOW OFTEN STUDIED 1865-1945 PERIOD: A LOT
002 186545-2 (2          ) 100                        HOW OFTEN STUDIED 1865-1945 PERIOD: SOME
003 186545-3 (3          ) 010                        HOW OFTEN STUDIED 1865-1945 PERIOD: NOT AT ALL
004 186545-? (M          ) 001                        HOW OFTEN STUDIED 1865-1945 PERIOD: MISSING
```

Table C-5 (continued)
Conditioning Variables Specific to the U.S. History Assessment

```
CONDITIONING VARIABLE ID:  SUBJ0028
DESCRIPTION:               HOW OFTEN HAVE YOU STUDIED THE POST 1945 PERIOD OF U.S. HISTORY?
GRADES/ASSESSMENTS:        N08, N12
CONDITIONING VAR LABEL:    POST1945
NAEP ID:                   H803604                    TOTAL NUMBER OF SPECIFIED CONTRASTS:    4
TYPE OF CONTRAST:          CLASS                      NUMBER OF INDEPENDENT CONTRASTS:        3

001 PS1945-1 (1          ) 000                        HOW OFTEN STUDIED POST 1945 PERIOD: A LOT
002 PS1945-2 (2          ) 100                        HOW OFTEN STUDIED POST 1945 PERIOD: SOME
003 PS1945-3 (3          ) 010                        HOW OFTEN STUDIED POST 1945 PERIOD: NOT AT ALL
004 PW1945-? (M          ) 001                        HOW OFTEN STUDIED POST 1945 PERIOD: MISSING

CONDITIONING VARIABLE ID:  SUBJ0029
DESCRIPTION:               ARE YOU CURRENTLY ENROLLED IN OR HAVE YOU TAKEN AN ADVANCED PLACEMENT COURSE IN
                           HISTORY?
GRADES/ASSESSMENTS:        N12
CONDITIONING VAR LABEL:    AP_HIST
NAEP ID:                   H803701                    TOTAL NUMBER OF SPECIFIED CONTRASTS:    3
TYPE OF CONTRAST:          CLASS                      NUMBER OF INDEPENDENT CONTRASTS:        2

001 APHIS-Y  (1          ) 0CAP HISTORY:  YES
002 APHIS-N  (2          ) 10AP HISTORY:  NO
003 APHIS-?  (M          ) 01AP HISTORY:  MISSING

CONDITIONING VARIABLE ID:  SUBJ0030
DESCRIPTION:               ABOUT HOW MANY QUESTIONS DID YOU GET RIGHT ON THE HISTORY TEST?
GRADES/ASSESSMENTS:        N04, N08, N12
CONDITIONING VAR LABEL:    #QUESTN+
NAEP ID:                   HM00101                    TOTAL NUMBER OF SPECIFIED CONTRASTS:    5
TYPE OF CONTRAST:          CLASS                      NUMBER OF INDEPENDENT CONTRASTS:        4

001 #QUEST+1 (1          ) 0000                       NUMBER QUESTIONS RIGHT:  ALMOST ALL
002 #QUEST+2 (2          ) 1000                       NUMBER QUESTIONS RIGHT:  MORE THAN HALF
003 #QUEST+3 (3          ) 0100                       NUMBER QUESTIONS RIGHT:  ABOUT HALF
004 #QUEST+4 (4          ) 0010                       NUMBER QUESTIONS RIGHT:  LESS THAN HALF
005 #QUEST+? (M          ) 0001                       NUMBER QUESTIONS RIGHT:  MISSING

CONDITIONING VARIABLE ID:  SUBJ0031
DESCRIPTION:               HOW HARD WAS THIS HISTORY TEST COMPARED TO OTHERS?
GRADES/ASSESSMENTS:        N04, N08, N12
CONDITIONING VAR LABEL:    TEST DIF
NAEP ID:                   HM00201                    TOTAL NUMBER OF SPECIFIED CONTRASTS:    5
TYPE OF CONTRAST:          CLASS                      NUMBER OF INDEPENDENT CONTRASTS:        4

001 TESTDIF1 (1          ) 0000                       TEST DIFFICULTY:  MUCH HARDER THAN OTHERS
002 TESTDIF2 (2          ) 1000                       TEST DIFFICULTY:  HARDER THAN OTHERS
003 TESTDIF3 (3          ) 0100                       TEST DIFFICULTY:  ABOUT AS HARD AS OTHERS
004 TESTDIF4 (4          ) 0010                       TEST DIFFICULTY:  EASIER THAN OTHERS
005 TESTDIF? (M          ) 0001                       TEST DIFFICULTY:  MISSING

CONDITIONING VARIABLE ID:  SUBJ0032
DESCRIPTION:               HOW HARD DID YOU TRY ON THIS TEST COMPARED TO OTHER HISTORY TESTS?
GRADES/ASSESSMENTS:        N04, N08, N12
CONDITIONING VAR LABEL:    TEST EFF
NAEP ID:                   HM00301                    TOTAL NUMBER OF SPECIFIED CONTRASTS:    5
TYPE OF CONTRAST:          CLASS                      NUMBER OF INDEPENDENT CONTRASTS:        4

001 TESTEFF1 (1          ) 0000                       TEST EFFORT:  MUCH HARDER THAN OTHERS
002 TESTEFF2 (2          ) 1000                       TEST EFFORT:  HARDER THAN OTHERS
003 TESTEFF3 (3          ) 0100                       TEST EFFORT:  ABOUT AS HARD AS OTHERS
004 TESTEFF4 (4          ) 0010                       TEST EFFORT:  NOT AS HARD AS OTHERS
005 TESTEFF? (M          ) 0001                       TEST EFFORT:  MISSING
```

Table C-5 (continued)

Conditioning Variables Specific to the U.S. History Assessment

```
CONDITIONING VARIABLE ID:  SUBJ0033
DESCRIPTION:               HOW IMPORTANT WAS IT TO YOU TO DO WELL ON THE HISTORY TEST?
GRADES/ASSESSMENTS:        N04, N08, N12
CONDITIONING VAR LABEL:    TEST IMP
NAEP ID:                   HM00401                           TOTAL NUMBER OF SPECIFIED CONTRASTS:      5
TYPE OF CONTRAST:          CLASS                             NUMBER OF INDEPENDENT CONTRASTS:          4

001 TESTIMP1 (1          ) 0000                              TEST IMPORTANCE:  VERY IMPORTANT
002 TESTIMP2 (2          ) 1000                              TEST IMPORTANCE:  IMPORTANT
003 TESTIMP3 (3          ) 0100                              TEST IMPORTANCE:  SOMEWHAT IMPORTANT
004 TESTIMP4 (4          ) 0010                              TEST IMPORTANCE:  NOT VERY IMPORTANT
005 TESTIMP? (M          ) 0001                              TEST IMPORTANCE:  MISSING

CONDITIONING VARIABLE ID:  SUBJ0034
DESCRIPTION:               HOW OFTEN WERE YOU ASKED TO WRITE LONG ANSWERS ON HISTORY TESTS?
GRADES/ASSESSMENTS:        N04, N08, N12
CONDITIONING VAR LABEL:    LONG ANS
NAEP ID:                   HM00501                           TOTAL NUMBER OF SPECIFIED CONTRASTS:      5
TYPE OF CONTRAST:          CLASS                             NUMBER OF INDEPENDENT CONTRASTS:          4

001 LNGANSW1 (1          ) 0000                              LONG ANSWERS:  AT LEAST ONCE A WEEK
002 LNGANSW2 (2          ) 1000                              LONG ANSWERS:  ONCE OR TWICE A MONTH
003 LNGANSW3 (3          ) 0100                              LONG ANSWERS:  ONCE OR TWICE A YEAR
004 LNGANSW4 (4          ) 0010                              LONG ANSWERS:  NEVER
005 LNGANSW? (M          ) 0001                              LONG ANSWERS:  MISSING

CONDITIONING VARIABLE ID:  TCHR0001
DESCRIPTION:               TEACHER MATCH STATUS WITH STUDENT
GRADES/ASSESSMENTS:        N04, N08
CONDITIONING VAR LABEL:    T_MATCH
NAEP ID:                   TCHMTCH                           TOTAL NUMBER OF SPECIFIED CONTRASTS:      3
TYPE OF CONTRAST:          CLASS                             NUMBER OF INDEPENDENT CONTRASTS:          2

001 TMCH-NO  (1,M        ) 00TEACHER MATCH:  NO MATCH
002 TMCH-PAR (2          ) 10TEACHER MATCH:  PARTIAL MATCH
003 TMCH-COM (3          ) 01TEACHER MATCH:  COMPLETE MATCH

CONDITIONING VARIABLE ID:  TCHR0002
DESCRIPTION:               TEACHER GENDER
GRADES/ASSESSMENTS:        N04, N08
CONDITIONING VAR LABEL:    T_GENDER
NAEP ID:                   T040001                           TOTAL NUMBER OF SPECIFIED CONTRASTS:      3
TYPE OF CONTRAST:          CLASS                             NUMBER OF INDEPENDENT CONTRASTS:          2

001 T_MALE   (1          ) 00TEACHER GENDER:  MALE
002 T_FEMALE (2          ) 10TEACHER GENDER:  FEMALE
003 T_SEX-?  (M          ) 01TEACHER GENDER:  MISSING, DOES NOT APPLY

CONDITIONING VARIABLE ID:  TCHR0003
DESCRIPTION:               TEACHER RACE/ETHNICITY
GRADES/ASSESSMENTS:        N04, N08
CONDITIONING VAR LABEL:    T_RACE
NAEP ID:                   T050801                           TOTAL NUMBER OF SPECIFIED CONTRASTS:      7
TYPE OF CONTRAST:          CLASS                             NUMBER OF INDEPENDENT CONTRASTS:          6

001 T_WHITE  (1          ) 000000                            TEACHER RACE/ETHNICITY:  WHITE
002 T_BLACK  (2          ) 100000                            TEACHER RACE/ETHNICITY:  BLACK
003 T_HISP   (3          ) 010000                            TEACHER RACE/ETHNICITY:  HISPANIC
004 T_ASIAN  (4          ) 001000                            TEACHER RACE/ETHNICITY:  ASIAN
005 T_PAC IS (5          ) 000100                            TEACHER RACE/ETHNICITY:  PACIFIC ISLANDER
```

Table C-5 (continued)

Conditioning Variables Specific to the U.S. History Assessment

```
006 T_AM IND (6          ) 000010                           TEACHER RACE/ETHNICITY:  AMERICAN INDIAN/ALASKAN
                                                            NATIV
007 T_RACE-? (M          ) 000001                           TEACHER RACE/ETHNICITY:  MISSING, DOES NOT APPLY

CONDITIONING VARIABLE ID:  TCHR0004
DESCRIPTION:               YEARS TEACHING ELEMENTARY/SECONDARY SCHOOL
GRADES/ASSESSMENTS:       N04, N08
CONDITIONING VAR LABEL:    T_YRSEXP
NAEP ID:                   T040301                          TOTAL NUMBER OF SPECIFIED CONTRASTS:     6
TYPE OF CONTRAST:          CLASS                            NUMBER OF INDEPENDENT CONTRASTS:         5

001 T_YREXP1 (1          ) 00000                            YEARS TEACHING:  2 OR LESS YEARS
002 T_YREXP2 (2          ) 10000                            YEARS TEACHING:  3-5 YEARS
003 T_YREXP3 (3          ) 01000                            YEARS TEACHING:  6-10 YEARS
004 T_YREXP4 (4          ) 00100                            YEARS TEACHING:  11-24 YEARS
005 T_YREXP5 (5          ) 00010                            YEARS TEACHING:  25 OR MORE YEARS
006 T_YREXP? (M          ) 00001                            YEARS TEACHING:  MISSING, DOES NOT APPLY

CONDITIONING VARIABLE ID:  TCHR0005
DESCRIPTION:               TEACHER GENERAL CERTIFICATION (ELEMENTARY, MIDDLE/JUNIOR, HIGH SCHOOL EDUCATION)
GRADES/ASSESSMENTS:       N04, N08
CONDITIONING VAR LABEL:    CERT GEN
NAEP ID:                   T040501                          TOTAL NUMBER OF SPECIFIED CONTRASTS:     4
TYPE OF CONTRAST:          CLASS                            NUMBER OF INDEPENDENT CONTRASTS:         3

001 CERTG-Y  (1          ) 000                              TEACHER GENERAL CERTIFICATION:  YES
002 CERTG-N  (2          ) 100                              TEACHER GENERAL CERTIFICATION:  NO
003 CERTG-NS (3          ) 010                              TEACHER GENERAL CERTIFICATION:  NOT OFFERED IN
                                                            STATE
004 CERTG-?  (M          ) 001                              TEACHER GENERAL CERTIFICATION:  MISSING, DOES NOT
                                                            APPLY

CONDITIONING VARIABLE ID:  TCHR0006
DESCRIPTION:               TEACHER CERTIFICATION IN READING
GRADES/ASSESSMENTS:       N04
CONDITIONING VAR LABEL:    CERT RED
NAEP ID:                   T040502                          TOTAL NUMBER OF SPECIFIED CONTRASTS:     4
TYPE OF CONTRAST:          CLASS                            NUMBER OF INDEPENDENT CONTRASTS:         3

001 CERTR-Y  (1          ) 000                              TEACHER READING CERTIFICATION:  YES
002 CERTR-N  (2          ) 100                              TEACHER READING CERTIFICATION:  NO
003 CERTR-NS (3          ) 010                              TEACHER READING CERTIFICATION:  NOT OFFERED IN
                                                            STATE
004 CERTR-?  (M          ) 001                              TEACHER READING CERTIFICATION:  MISSING, DOES NOT
                                                            APPLY

CONDITIONING VARIABLE ID:  TCHR0007
DESCRIPTION:               TEACHER CERTIFICATION MIDDLE/JUNIOR HIGH SCHOOL/SECONDARY ENGLISH/LANGUAGE ARTS
GRADES/ASSESSMENTS:       N04
CONDITIONING VAR LABEL:    CERT LAN
NAEP ID:                   T040508                          TOTAL NUMBER OF SPECIFIED CONTRASTS:     4
TYPE OF CONTRAST:          CLASS                            NUMBER OF INDEPENDENT CONTRASTS:         3

001 CERTL-Y  (1          ) 000                              TEACHER ENGLISH/LANGUAGE ARTS CERTIFICATION: YES
002 CERTL-N  (2          ) 100                              TEACHER ENGLISH/LANGUAGE ARTS CERTIFICATION: NO
003 CERTL-NS (3          ) 010                              TEACHER ENGLISH/LANGUAGE ARTS CERTIFICATION: NOT
                                                            OFFERED IN STATE
004 CERTL-?  (M          ) 001                              TEACHER ENGLISH/LANGUAGE ARTS CERTIFICATION:
                                                            MISSING, DOES NOT APPLY

CONDITIONING VARIABLE ID:  TCHR0008
DESCRIPTION:               YEARS TEACHING SOCIAL STUDIES, HISTORY, OR GEOGRAPHY
```

Table C-5 (continued)

Conditioning Variables Specific to the U.S. History Assessment

```
GRADES/ASSESSMENTS:         N08
CONDITIONING VAR LABEL:     T_YRSEXP
NAEP ID:                    T054601                    TOTAL NUMBER OF SPECIFIED CONTRASTS:     6
TYPE OF CONTRAST:           CLASS                      NUMBER OF INDEPENDENT CONTRASTS:         5

001 T_YREXP1 (1          ) 00000                       YEARS TEACHING S.S., HIST. OR GEOG:  2 OR LESS
                                                       YEARS
002 T_YREXP2 (2          ) 10000                       YEARS TEACHING S.S., HIST. OR GEOG:  3-5 YEARS
003 T_YREXP3 (3          ) 01000                       YEARS TEACHING S.S., HIST. OR GEOG:  6-10 YEARS
004 T_YREXP4 (4          ) 00100                       YEARS TEACHING S.S., HIST. OR GEOG:  11-24 YEARS
005 T_YREXP5 (5          ) 00010                       YEARS TEACHING S.S., HIST. OR GEOG:  25 OR MORE
                                                       YEARS
006 T_YREXP? (M          ) 00001                       YEARS TEACHING S.S., HIST. OR GEOG:  MISSING, DOES
                                                       NOT APPLY

CONDITIONING VARIABLE ID:   TCHR0009
DESCRIPTION:                HOW WELL DOES YOUR SCHOOL PROVIDE YOU WITH INSTRUCTIONAL MATERIAL/RESOURCES YOU
                            NEED?
GRADES/ASSESSMENTS:         N04
CONDITIONING VAR LABEL:     RESOURCE
NAEP ID:                    T041201                    TOTAL NUMBER OF SPECIFIED CONTRASTS:     5
TYPE OF CONTRAST:           CLASS                      NUMBER OF INDEPENDENT CONTRASTS:         4

001 RESOURC1 (1          ) 0000                        RESOURCES:  GET ALL
002 RESOURC2 (2          ) 1000                        RESOURCES:  GET MOST
003 RESOURC3 (3          ) 0100                        RESOURCES:  GET SOME
004 RESOURC4 (4          ) 0010                        RESOURCES:  DON'T GET
005 RESOURC? (M          ) 0001                        RESOURCES:  MISSING, DOES NOT APPLY

CONDITIONING VARIABLE ID:   TCHR0010
DESCRIPTION:                TEACHER UNDERGRADUATE MAJOR IN EDUCATION
GRADES/ASSESSMENTS:         N04, N08
CONDITIONING VAR LABEL:     UGRAD ED
NAEP ID:                    T040701                    TOTAL NUMBER OF SPECIFIED CONTRASTS:     2
TYPE OF CONTRAST:           CLASS                      NUMBER OF INDEPENDENT CONTRASTS:         1

001 UGR ED-? (0,M        ) 0 TEACHER UNDERGRAD EDUCATION MAJOR:  MISSING, DOES NOT APPLY
002 UGR ED-Y (1          ) 1 TEACHER UNDERGRAD EDUCATION MAJOR:  YES

CONDITIONING VARIABLE ID:   TCHR0011
DESCRIPTION:                TEACHER GRADUATE MAJOR IN EDUCATION
GRADES/ASSESSMENTS:         N04, N08
CONDITIONING VAR LABEL:     GRAD ED
NAEP ID:                    T040801                    TOTAL NUMBER OF SPECIFIED CONTRASTS:     2
TYPE OF CONTRAST:           CLASS                      NUMBER OF INDEPENDENT CONTRASTS:         1

001 GRA ED-? (0,M        ) 0 TEACHER GRADUATE EDUCATION MAJOR:  MISSING, DOES NOT APPLY
002 GRA ED-Y (1          ) 1 TEACHER GRADUATE EDUCATION MAJOR:  YES

CONDITIONING VARIABLE ID:   TCHR0012
DESCRIPTION:                NO TEACHER GRADUATE-LEVEL STUDY
GRADES/ASSESSMENTS:         N04, N08
CONDITIONING VAR LABEL:     NO GRAD
NAEP ID:                    T040806                    TOTAL NUMBER OF SPECIFIED CONTRASTS:     2
TYPE OF CONTRAST:           CLASS                      NUMBER OF INDEPENDENT CONTRASTS:         1

001 NOGRAD-? (0,M        ) 0 NO TEACHER GRADUATE STUDY:  MISSING, DOES NOT APPLY
002 NOGRAD-Y (1          ) 1 NO TEACHER GRADUATE STUDY:  YES

CONDITIONING VARIABLE ID:   TCHR0013
DESCRIPTION:                HOW MANY YEARS IN TOTAL HAVE YOU TAUGHT READING? (4TH GRADE)
GRADES/ASSESSMENTS:         N04
```

Table C-5 (continued)

Conditioning Variables Specific to the U.S. History Assessment

```
CONDITIONING VAR LABEL:     T4REDYRS
NAEP ID:                    T049901                  TOTAL NUMBER OF SPECIFIED CONTRASTS:     6
TYPE OF CONTRAST:           CLASS                    NUMBER OF INDEPENDENT CONTRASTS:         5

001 T4REDYR1 (1          ) 00000                     YEARS TEACHING READING:  2 YEARS OR LESS
002 T4REDYR2 (2          ) 10000                     YEARS TEACHING READING:  3-5 YEARS
003 T4REDYR3 (3          ) 01000                     YEARS TEACHING READING:  6-10 YEARS
004 T4REDYR4 (4          ) 00100                     YEARS TEACHING READING:  11-24 YEARS
005 T4REDYR5 (5          ) 00010                     YEARS TEACHING READING:  25 YEARS OR MORE
006 T4REDYR? (M          ) 00001                     YEARS TEACHING READING:  MISSING, DOES NOT APPLY

CONDITIONING VARIABLE ID:  TCHR0014
DESCRIPTION:               DURING THE LAST YEAR IN SOCIAL STUDIES STAFF DEVELOPMENT WORKSHOPS  OR SEMINARS?
GRADES/ASSESSMENTS:        N04, N08
CONDITIONING VAR LABEL:    TCHWRKSH
NAEP ID:                   T050301                   TOTAL NUMBER OF SPECIFIED CONTRASTS:     6
TYPE OF CONTRAST:          CLASS                     NUMBER OF INDEPENDENT CONTRASTS:         5

001 TWKSP-0  (1          ) 00000                     TEACHER WORKSHOPS OR SEMINARS:  NONE
002 TWKSP-<6 (2          ) 10000                     TEACHER WORKSHOPS OR SEMINARS:  LESS THAN 6 HOURS
003 TWKSP-15 (3          ) 01000                     TEACHER WORKSHOPS OR SEMINARS:  6 - 15 HOURS
004 TWKSP-35 (4          ) 00100                     TEACHER WORKSHOPS OR SEMINARS:  16 - 35 HOURS
005 TWKSP>35 (5          ) 00010                     TEACHER WORKSHOPS OR SEMINARS:  MORE THAN 35 HOURS
006 TWKSP-?  (M          ) 00001                     TEACHER WORKSHOPS OR SEMINARS:  MISSING, DOES NOT
                                                     APPLY

CONDITIONING VARIABLE ID:  TCHR0015
DESCRIPTION:               WHAT TYPE OF TEACHING CERTIFICATION DO YOU HAVE THAT IS RECOGNIZED BY THE STATE IN
                           WHICH YOU TEACH?
GRADES/ASSESSMENTS:        N04, N08
CONDITIONING VAR LABEL:    TYP CERT
NAEP ID:                   T050001                   TOTAL NUMBER OF SPECIFIED CONTRASTS:     5
TYPE OF CONTRAST:          CLASS                     NUMBER OF INDEPENDENT CONTRASTS:         4

001 CERT-NO  (1          ) 0000                      TYPE OF TEACHING CERTIFICATION:  NONE
002 CERT-TMP (2          ) 1000                      TYPE OF TEACHING CERTIFICATION:  TEMPORARY,
                                                     PROBATIONAL, PROVISIONAL, EMERGENCY
003 CERT-REG (3          ) 0100                      TYPE OF TEACHING CERTIFICATION:  REGULAR, BUT NOT
                                                     HIGHEST
004 CERT-HGH (4          ) 0010                      TYPE OF TEACHING CERTIFICATION:  HIGHEST AVAILABLE
005 CERT-?   (M          ) 0001                      TYPE OF TEACHING CERTIFICATION:  MISSING, DOES NOT
                                                     APPLY

CONDITIONING VARIABLE ID:  TCHR0016
DESCRIPTION:               WHAT IS THE HIGHEST ACADEMIC DEGREE YOU HOLD?
GRADES/ASSESSMENTS:        N04, N08
CONDITIONING VAR LABEL:    T_DEGREE
NAEP ID:                   T050101                   TOTAL NUMBER OF SPECIFIED CONTRASTS:     8
TYPE OF CONTRAST:          CLASS                     NUMBER OF INDEPENDENT CONTRASTS:         7

001 TDEG-HSD (1          ) 0000000                   TEACHER HIGHEST DEGREE:  HIGH SCHOOL DIPLOMA
002 TDEG-ASC (2          ) 1000000                   TEACHER HIGHEST DEGREE:  ASSOCIATES/VOCATIONAL
003 TDEG-BAC (3          ) 0100000                   TEACHER HIGHEST DEGREE:  BACHELOR'S HIGHEST DEGREE
004 TDEG-MAS (4          ) 0010000                   TEACHER HIGHEST DEGREE:  MASTER'S HIGHEST DEGREE
005 TDEG-EDS (5          ) 0001000                   TEACHER HIGHEST DEGREE:  EDUCATION SPECIALIST
006 TDEG-DOC (6          ) 0000100                   TEACHER HIGHEST DEGREE:  DOCTORATE
007 TDEG-PRO (7          ) 0000010                   TEACHER HIGHEST DEGREE:  PROFESSIONAL HIGHEST
                                                     DEGREE
008 TDEG-?   (M          ) 0000001                   TEACHER HIGHEST DEGREE:  MISSING, DOES NOT APPLY

CONDITIONING VARIABLE ID:  TCHR0017
DESCRIPTION:               TEACHER ENGLISH UNDERGRADUATE MAJOR
```

Table C-5 (continued)

Conditioning Variables Specific to the U.S. History Assessment

```
GRADES/ASSESSMENTS:        N04
CONDITIONING VAR LABEL:    UGRD ENG
NAEP ID:                   T040706                          TOTAL NUMBER OF SPECIFIED CONTRASTS:      2
TYPE OF CONTRAST:          CLASS                            NUMBER OF INDEPENDENT CONTRASTS:          1

001 UGRENG-Y (1          ) 0 TEACHER ENGLISH UNDERGRADUATE MAJOR:  YES
002 UGRENG-? (0,M        ) 1 TEACHER ENGLISH UNDERGRADUATE MAJOR:  MISSING, DOES NOT APPLY

CONDITIONING VARIABLE ID:  TCHR0018
DESCRIPTION:               TEACHER READING/LANGUAGE ARTS UNDERGRADUATE MAJOR
GRADES/ASSESSMENTS:        N04
CONDITIONING VAR LABEL:    UGRD RED
NAEP ID:                   T040707                          TOTAL NUMBER OF SPECIFIED CONTRASTS:      2
TYPE OF CONTRAST:          CLASS                            NUMBER OF INDEPENDENT CONTRASTS:          1

001 UGRRED-Y (1          ) 0 TEACHER READING UNDERGRADUATE MAJOR:  YES
002 UGRRED-? (0,M        ) 1 TEACHER READING UNDERGRADUATE MAJOR:  MISSING, DOES NOT APPLY

CONDITIONING VARIABLE ID:  TCHR0019
DESCRIPTION:               TEACHER ENGLISH GRADUATE MAJOR
GRADES/ASSESSMENTS:        N04
CONDITIONING VAR LABEL:    GRAD ENG
NAEP ID:                   T040807                          TOTAL NUMBER OF SPECIFIED CONTRASTS:      2
TYPE OF CONTRAST:          CLASS                            NUMBER OF INDEPENDENT CONTRASTS:          1

001 GRDENG-Y (1          ) 0 TEACHER ENGLISH GRADUATE MAJOR:       YES
002 GRDENG-? (0,M        ) 1 TEACHER ENGLISH GRADUATE MAJOR:       MISSING, DOES NOT APPLY

CONDITIONING VARIABLE ID:  TCHR0020
DESCRIPTION:               TEACHER READING/LANGUAGE ARTS GRADUATE MAJOR
GRADES/ASSESSMENTS:        N04
CONDITIONING VAR LABEL:    GRAD RED
NAEP ID:                   T040808                          TOTAL NUMBER OF SPECIFIED CONTRASTS:      2
TYPE OF CONTRAST:          CLASS                            NUMBER OF INDEPENDENT CONTRASTS:          1

001 GRDRED-Y (1          ) 0 TEACHER READING GRADUATE MAJOR:       YES
002 GRDRED-? (0,M        ) 1 TEACHER READING GRADUATE MAJOR:       MISSING, DOES NOT APPLY

CONDITIONING VARIABLE ID:  TCHR0021
DESCRIPTION:               HOW MUCH TIME HAVE YOU SPENT LAST YEAR IN READING DEVELOPMENT WORKSHOPS/SEMINARS?
GRADES/ASSESSMENTS:        N04
CONDITIONING VAR LABEL:    READ DEV
NAEP ID:                   T050201                          TOTAL NUMBER OF SPECIFIED CONTRASTS:      6
TYPE OF CONTRAST:          CLASS                            NUMBER OF INDEPENDENT CONTRASTS:          5

001 READDEV1 (1          ) 00000                            TIME IN READING DEVELOPMENT WORKSHOPS:  NONE
002 READDEV2 (2          ) 10000                            TIME IN READING DEVELOPMENT WORKSHOPS:  LESS THAN
                                                            6 HOURS
003 READDEV3 (3          ) 01000                            TIME IN READING DEVELOPMENT WORKSHOPS:  6-15 HOURS
004 READDEV4 (4          ) 00100                            TIME IN READING DEVELOPMENT WORKSHOPS:  16-35
                                                            HOURS
005 READDEV5 (5          ) 00010                            TIME IN READING DEVELOPMENT WORKSHOPS:  MORE THAN
                                                            35 HOURS
006 READDEV5 (M          ) 00001                            TIME IN READING DEVELOPMENT WORKSHOPS:  MISSING,
                                                            DOES NOT APPLY

CONDITIONING VARIABLE ID:  TCHR0022
DESCRIPTION:               HOW MANY HOURS DO YOU HAVE DESIGNATED AS PREPERATION PERIODS PER WEEK?
GRADES/ASSESSMENTS:        N04, N08
CONDITIONING VAR LABEL:    PREP PER
NAEP ID:                   T051101                          TOTAL NUMBER OF SPECIFIED CONTRASTS:      5
TYPE OF CONTRAST:          CLASS                            NUMBER OF INDEPENDENT CONTRASTS:          4
```

Table C-5 (continued)

Conditioning Variables Specific to the U.S. History Assessment

```
001 PRPER-0  (1          ) 0000                    TEACHER WEEKLY PREPARTION PERIODS:  NONE
002 PRPER-<1 (2          ) 1000                    TEACHER WEEKLY PREPARTION PERIODS:  LESS THAN 1
003 PRPER-12 (3          ) 0100                    TEACHER WEEKLY PREPARTION PERIODS:  1 TO 2
004 PRPER->2 (4          ) 0010                    TEACHER WEEKLY PREPARTION PERIODS:  MORE THAN 2
005 PRPER-?  (M          ) 0001                    TEACHER WEEKLY PREPARTION PERIODS:  MISSING, DOES
                                                   NOT APPLY

CONDITIONING VARIABLE ID:  TCHR0023
DESCRIPTION:               ARE CURRICULUM SPECIALISTS AVAILABLE FOR READING?
GRADES/ASSESSMENTS:        N04
CONDITIONING VAR LABEL:    CURSPECS
NAEP ID:                   T041301                 TOTAL NUMBER OF SPECIFIED CONTRASTS:     3
TYPE OF CONTRAST:          CLASS                   NUMBER OF INDEPENDENT CONTRASTS:         2

001 CSPECS-Y (1          ) 00READING CURRICULUM SPECIALISTS:  YES
002 CSPECS-N (2          ) 10READING CURRICULUM SPECIALISTS:  NO
003 CSPECS-? (M          ) 01READING CURRICULUM SPECIALISTS:  MISSING, DOES NOT APPLY

CONDITIONING VARIABLE ID:  TCHR0024
DESCRIPTION:               TEACHER CERTIFICATION IN HISTORY
GRADES/ASSESSMENTS:        N04, N08
CONDITIONING VAR LABEL:    CERT HIS
NAEP ID:                   T040507                 TOTAL NUMBER OF SPECIFIED CONTRASTS:     4
TYPE OF CONTRAST:          CLASS                   NUMBER OF INDEPENDENT CONTRASTS:         3

001 CERTH-Y  (1          ) 000                     TEACHER HISTORY CERTIFICATION:  YES
002 CERTH-N  (2          ) 100                     TEACHER HISTORY CERTIFICATION:  NO
003 CERTH-NS (3          ) 010                     TEACHER HISTORY CERTIFICATION:  NOT OFFERED IN
                                                   STATE
004 CERTH-?  (M          ) 001                     TEACHER HISTORY CERTIFICATION:  MISSING, DOES NOT
                                                   APPLY

CONDITIONING VARIABLE ID:  TCHR0025
DESCRIPTION:               TEACHER CERTIFICATION IN GEOGRAPHY
GRADES/ASSESSMENTS:        N04, N08
CONDITIONING VAR LABEL:    CERT GEO
NAEP ID:                   T040506                 TOTAL NUMBER OF SPECIFIED CONTRASTS:     4
TYPE OF CONTRAST:          CLASS                   NUMBER OF INDEPENDENT CONTRASTS:         3

001 CERTG-Y  (1          ) 000                     TEACHER GEOGRAPHY CERTIFICATION: YES
002 CERTG-N  (2          ) 100                     TEACHER GEOGRAPHY CERTIFICATION: NO
003 CERTG-NS (3          ) 010                     TEACHER GEOGRAPHY CERTIFICATION: NOT OFFERED IN
                                                   STATE
004 CERTG-?  (M          ) 001                     TEACHER GEOGRAPHY CERTIFICATION: MISSING, DOES NOT
                                                   APPLY

CONDITIONING VARIABLE ID:  TCHR0026
DESCRIPTION:               TEACHER CERTIFICATION IN MIDDLE/JR. H.S. SOCIAL STUDIES
GRADES/ASSESSMENTS:        N08
CONDITIONING VAR LABEL:    CERT MSS
NAEP ID:                   T040509                 TOTAL NUMBER OF SPECIFIED CONTRASTS:     4
TYPE OF CONTRAST:          CLASS                   NUMBER OF INDEPENDENT CONTRASTS:         3

001 CERTM-Y  (1          ) 000                     TEACHER MIDDLE/JR. H.S. SOCIAL STUDIES
                                                   CERTIFICATION: YES
002 CERTM-N  (2          ) 100                     TEACHER MIDDLE/JR. H.S. SOCIAL STUDIES
                                                   CERTIFICATION: NO
003 CERTM-NS (3          ) 010                     TEACHER MIDDLE/JR. H.S. SOCIAL STUDIES
                                                   CERTIFICATION: NOT OFFERED IN STATE
004 CERTM-?  (M          ) 001                     TEACHER MIDDLE/JR. H.S. SOCIAL STUDIES
                                                   CERTIFICATION: MISSING, DOES NOT APPLY
```

BEST COPY AVAILABLE

Table C-5 (continued)

Conditioning Variables Specific to the U.S. History Assessment

```
CONDITIONING VARIABLE ID:  TCHR0027
DESCRIPTION:               TEACHER CERTIFICATION IN SECONDARY SOCIAL STUDIES TEACHER
GRADES/ASSESSMENTS:        N08
CONDITIONING VAR LABEL:    CERT_SSS
NAEP ID:                   T040510                           TOTAL NUMBER OF SPECIFIED CONTRASTS:      4
TYPE OF CONTRAST:          CLASS                             NUMBER OF INDEPENDENT CONTRASTS:          3

001 CERTS-Y  (1          ) 000                               TEACHER SECONDARY SOCIAL STUDIES TEACHER
                                                             CERTIFICATION: YES
002 CERTS-N  (2          ) 100                               TEACHER SECONDARY SOCIAL STUDIES TEACHER
                                                             CERTIFICATION: NO
003 CERTS-NS (3          ) 010                               TEACHER SECONDARY SOCIAL STUDIES TEACHER
                                                             CERTIFICATION: NOT OFFERED IN STATE
004 CERTS-?  (M          ) 001                               TEACHER SECONDARY SOCIAL STUDIES TEACHER
                                                             CERTIFICATION: MISSING, DOES NOT APPLY

CONDITIONING VARIABLE ID:  TCHR0028
DESCRIPTION:               TEACHER CERTIFICATION IN OTHER
GRADES/ASSESSMENTS:        N04, N08
CONDITIONING VAR LABEL:    CERT OTH
NAEP ID:                   T040505                           TOTAL NUMBER OF SPECIFIED CONTRASTS:      4
TYPE OF CONTRAST:          CLASS                             NUMBER OF INDEPENDENT CONTRASTS:          3

001 CERTO-Y  (1          ) 000                               TEACHER OTHER CERTIFICATION:    YES
002 CERTO-N  (2          ) 100                               TEACHER OTHER CERTIFICATION:    NO
003 CERTO-NS (3          ) 010                               TEACHER OTHER CERTIFICATION:    NOT OFFERED IN
                                                             STATE
004 CERTO-?  (M          ) 001                               TEACHER OTHER CERTIFICATION:    MISSING, DOES NOT
                                                             APPLY

CONDITIONING VARIABLE ID:  TCHR0029
DESCRIPTION:               TEACHER UNDERGRADUATE MAJOR IN GEOGRAPHY
GRADES/ASSESSMENTS:        N04, N08
CONDITIONING VAR LABEL:    UGRD GEO
NAEP ID:                   T040708                           TOTAL NUMBER OF SPECIFIED CONTRASTS:      2
TYPE OF CONTRAST:          CLASS                             NUMBER OF INDEPENDENT CONTRASTS:          1

001 UGR GE-? (0,M        ) 0 TEACHER UNDERGRAD GEOGRAPHY MAJOR:  MISSING, DOES NOT APPLY
002 UGR GE-Y (1          ) 1 TEACHER UNDERGRAD GEOGRAPHY MAJOR:  YES

CONDITIONING VARIABLE ID:  TCHR0030
DESCRIPTION:               TEACHER UNDERGRADUATE MAJOR IN HISTORY
GRADES/ASSESSMENTS:        N04, N08
CONDITIONING VAR LABEL:    UGRD HIS
NAEP ID:                   T040709                           TOTAL NUMBER OF SPECIFIED CONTRASTS:      2
TYPE OF CONTRAST:          CLASS                             NUMBER OF INDEPENDENT CONTRASTS:          1

001 UGR HI-? (0,M        ) 0 TEACHER UNDERGRAD HISTORY MAJOR:    MISSING, DOES NOT APPLY
002 UGR HI-Y (1          ) 1 TEACHER UNDERGRAD HISTORY MAJOR:    YES

CONDITIONING VARIABLE ID:  TCHR0031
DESCRIPTION:               TEACHER UNDERGRADUATE MAJOR IN SOCIAL STUDIES EDUCATION
GRADES/ASSESSMENTS:        N04, N08
CONDITIONING VAR LABEL:    UGRD SS
NAEP ID:                   T040710                           TOTAL NUMBER OF SPECIFIED CONTRASTS:      2
TYPE OF CONTRAST:          CLASS                             NUMBER OF INDEPENDENT CONTRASTS:          1

001 UGR SS-? (0,M        ) 0 TEACHER UNDERGRAD SOCIAL STUDIES EDUCATION MAJOR: MISSING, DOES NOT APPLY
002 UGR SS-Y (1          ) 1 TEACHER UNDERGRAD SOCIAL STUDIES EDUCATION MAJOR: YES

CONDITIONING VARIABLE ID:  TCHR0032
DESCRIPTION:               TEACHER UNDERGRADUATE MAJOR IN OTHER
```

Table C-5 (continued)

Conditioning Variables Specific to the U.S. History Assessment

```
GRADES/ASSESSMENTS:        N04, N08
CONDITIONING VAR LABEL:    UGRD OTH
NAEP ID:                   T040705                          TOTAL NUMBER OF SPECIFIED CONTRASTS:     2
TYPE OF CONTRAST:          CLASS                            NUMBER OF INDEPENDENT CONTRASTS:         1

001 UGR OT-? (0,M        ) 0 TEACHER UNDERGRAD OTHER MAJOR:      MISSING, DOES NOT APPLY
002 UGR OT-Y (1          ) 1 TEACHER UNDERGRAD OTHER MAJOR:      YES

CONDITIONING VARIABLE ID:  TCHR0033
DESCRIPTION:               TEACHER GRADUATE MAJOR IN GEOGRAPHY
GRADES/ASSESSMENTS:        N04, N08
CONDITIONING VAR LABEL:    GRD GEO
NAEP ID:                   T040809                          TOTAL NUMBER OF SPECIFIED CONTRASTS:     2
TYPE OF CONTRAST:          CLASS                            NUMBER OF INDEPENDENT CONTRASTS:         1

001 GRD GE-? (0,M        ) 0 TEACHER GRADUATE GEOGRAPHY MAJOR:  MISSING, DOES NOT APPLY
002 GRD GE-Y (1          ) 1 TEACHER GRADUATE GEOGRAPHY MAJOR:  YES

CONDITIONING VARIABLE ID:  TCHR0034
DESCRIPTION:               TEACHER GRADUATE MAJOR IN HISTORY
GRADES/ASSESSMENTS:        N04, N08
CONDITIONING VAR LABEL:    GRD HIS
NAEP ID:                   T040810                          TOTAL NUMBER OF SPECIFIED CONTRASTS:     2
TYPE OF CONTRAST:          CLASS                            NUMBER OF INDEPENDENT CONTRASTS:         1

001 GRD HI-? (0,M        ) 0 TEACHER GRADUATE HISTORY MAJOR:    MISSING, DOES NOT APPLY
002 GRD HI-Y (1          ) 1 TEACHER GRADUATE HISTORY MAJOR:    YES

CONDITIONING VARIABLE ID:  TCHR0035
DESCRIPTION:               TEACHER GRADUATE MAJOR IN SOCIAL STUDIES EDUCATION
GRADES/ASSESSMENTS:        N04, N08
CONDITIONING VAR LABEL:    GRD SS
NAEP ID:                   T040811                          TOTAL NUMBER OF SPECIFIED CONTRASTS:     2
TYPE OF CONTRAST:          CLASS                            NUMBER OF INDEPENDENT CONTRASTS:         1

001 GRD SS-? (0,M        ) 0 TEACHER GRADUATE SOCIAL STUDIES EDUCATION MAJOR: MISSING, DOES NOT APPLY
002 GRD SS-Y (1          ) 1 TEACHER GRADUATE SOCIAL STUDIES EDUCATION MAJOR: YES

CONDITIONING VARIABLE ID:  TChR0036
DESCRIPTION:               TEACHER GRADUATE MAJOR -OTHER
GRADES/ASSESSMENTS:        N04, N08
CONDITIONING VAR LABEL:    GRD OTH
NAEP ID:                   T040805                          TOTAL NUMBER OF SPECIFIED CONTRASTS:     2
TYPE OF CONTRAST:          CLASS                            NUMBER OF INDEPENDENT CONTRASTS:         1

001 GRA OT-? (0,M        ) 0 TEACHER GRADUATE OTHER MAJOR:  MISSING, DOES NOT APPLY
002 GRA OT-Y (1          ) 1 TEACHER GRADUATE OTHER MAJOR:  YES

CONDITIONING VARIABLE ID:  TCHR0037
DESCRIPTION:               DO YOU SAVE YOUR STUDENTS' SOCIAL STUDIES, HISTORY OR GEOGRAPHY WORK IN FOLDERS OR
                           PORTFOLIOS?
GRADES/ASSESSMENTS:        N04, N08
CONDITIONING VAR LABEL:    PORT-SS
NAEP ID:                   T050501                          TOTAL NUMBER OF SPECIFIED CONTRASTS:     3
TYPE OF CONTRAST:          CLASS                            NUMBER OF INDEPENDENT CONTRASTS:         2

001 PORT---Y (1          ) 00PORTFOLIOS: YES
002 PORT---N (2          ) 10PORTFOLIOS: NO
003 PORT---? (M          ) 01PORTFOLIOS: MISSING

CONDITIONING VARIABLE ID:  TCHR0038
DESCRIPTION:               DO YOU SAVE YOUR STUDENTS' READING WORK IN FOLDERS OR PORTFOLIOS?
```

Table C-5 (continued)

Conditioning Variables Specific to the U.S. History Assessment

```
GRADES/ASSESSMENTS:        N04
CONDITIONING VAR LABEL:    PORT-RED
NAEP ID:                   T050601                   TOTAL NUMBER OF SPECIFIED CONTRASTS:     3
TYPE OF CONTRAST:          CLASS                     NUMBER OF INDEPENDENT CONTRASTS:         2

001 PORTR--Y (1          ) 00READING PORTFOLIOS: YES
002 PORTR--N (2          ) 10READING PORTFOLIOS: NO
003 PORTR--? (M          ) 01READING PORTFOLIOS: MISSING

CONDITIONING VARIABLE ID:  TCHR0039
DESCRIPTION:               HOW MANY YEARS IN TOTAL HAVE YOU TAUGHT HISTORY.GEPGRAPHY, OR SOCIAL STUDIES? (4TH
                           GRADE)
GRADES/ASSESSMENTS:        N04
CONDITIONING VAR LABEL:    T4HISYRS
NAEP ID:                   T049902                   TOTAL NUMBER OF SPECIFIED CONTRASTS:     6
TYPE OF CONTRAST:          CLASS                     NUMBER OF INDEPENDENT CONTRASTS:         5

001 T4HISYR1 (1          ) 00000                     YEARS TEACHING HISTORY.GEPGRAPHY, OR SOCIAL
                                                     STUDIES: 2 YEARS OR LESS
002 T4HISYR2 (2          ) 10000                     YEARS TEACHING HISTORY.GEPGRAPHY, OR SOCIAL
                                                     STUDIES: 3-5 YEARS
003 T4HISYR3 (3          ) 01000                     YEARS TEACHING HISTORY.GEPGRAPHY, OR SOCIAL
                                                     STUDIES: 6-10 YEARS
004 T4HISYR4 (4          ) 00100                     YEARS TEACHING HISTORY.GEPGRAPHY, OR SOCIAL
                                                     STUDIES: 11-24 YEARS
005 T4HISYR5 (5          ) 00010                     YEARS TEACHING HISTORY.GEPGRAPHY, OR SOCIAL
                                                     STUDIES: 25 YEARS OR MORE
006 T4HISYR? (M          ) 00001                     YEARS TEACHING HISTORY.GEPGRAPHY, OR SOCIAL
                                                     STUDIES: MISSING, DOES NOT APPLY

CONDITIONING VARIABLE ID:  TCHR0040
DESCRIPTION:               DURING THE LAST YEAR IN GEOGRAPHY STAFF DEVELOPEMENT WORKSHOPS OR SEMINARS?
GRADES/ASSESSMENTS:        N04
CONDITIONING VAR LABEL:    TCHWRKSG
NAEP ID:                   T050401                   TOTAL NUMBER OF SPECIFIED CONTRASTS:     6
TYPE OF CONTRAST:          CLASS                     NUMBER OF INDEPENDENT CONTRASTS:         5

001 TWKSG-0  (1          ) 00000                     TEACHER WORKSHOPS OR SEMINARS IN GEOGRAPHY:  NONE
002 TWKSG-<6 (2          ) 10000                     TEACHER WORKSHOPS OR SEMINARS IN GEOGRAPHY:  LESS
                                                     THAN 6 HOURS
003 TWKSG-15 (3          ) 01000                     TEACHER WORKSHOPS OR SEMINARS IN GEOGRAPHY:  6 -
                                                     15 HOURS
004 TWKSG-35 (4          ) 00100                     TEACHER WORKSHOPS OR SEMINARS IN GEOGRAPHY:  16 -
                                                     35 HOURS
005 TWKSG>35 (5          ) 00010                     TEACHER WORKSHOPS OR SEMINARS IN GEOGRAPHY:  MORE
                                                     THAN 35 HOURS
006 TWKSG-?  (M          ) 00001                     TEACHER WORKSHOPS OR SEMINARS IN GEOGRAPHY:
                                                     MISSING, DOES NOT APPLY

CONDITIONING VARIABLE ID:  TCHR0041
DESCRIPTION:               ARE CURRICULUM SPECIALISTS AVAILABLE FOR GEOGRAPHY?
GRADES/ASSESSMENTS:        N04
CONDITIONING VAR LABEL:    CURSPECG
NAEP ID:                   T041303                   TOTAL NUMBER OF SPECIFIED CONTRASTS:     3
TYPE OF CONTRAST:          CLASS                     NUMBER OF INDEPENDENT CONTRASTS:         2

001 CSPECG-Y (1          ) 00GEOGRAPHY CURRICULUM SPECIALISTS: YES
002 CSPECG-N (2          ) 10GEOGRAPHY CURRICULUM SPECIALISTS: NO
003 CSPECG-? (M          ) 01GEOGRAPHY CURRICULUM SPECIALISTS: MISSING, DOES NOT APPLY

CONDITIONING VARIABLE ID:  TCHR0042
DESCRIPTION:               ARE CURRICULUM SPECIALISTS AVAILABLE FOR SOCIAL STUDIES?
```

Table C-5 (continued)
Conditioning Variables Specific to the U.S. History Assessment

```
GRADES/ASSESSMENTS:         N04
CONDITIONING VAR LABEL:     CURSPECS
NAEP ID:                    T041304                    TOTAL NUMBER OF SPECIFIED CONTRASTS:     3
TYPE OF CONTRAST:           CLASS                      NUMBER OF INDEPENDENT CONTRASTS:         2

001 CSPECS-Y (1          ) 00SOCIAL STUDIES CURRICULUM SPECIALISTS: YES
002 CSPECS-N (2          ) 10SOCIAL STUDIES CURRICULUM SPECIALISTS: NO
003 CSPECS-? (M          ) 01SOCIAL STUDIES CURRICULUM SPECIALISTS: MISSING, DOES NOT APPLY

CONDITIONING VARIABLE ID:  TCHR0043
DESCRIPTION:                ARE CURRICULUM SPECIALISTS AVAILABLE FOR HISTORY?
GRADES/ASSESSMENTS:         N04
CONDITIONING VAR LABEL:     CURSPECH
NAEP ID:                    T041305                    TOTAL NUMBER OF SPECIFIED CONTRASTS:     3
TYPE OF CONTRAST:           CLASS                      NUMBER OF INDEPENDENT CONTRASTS:         2

001 CSPECH-Y (1          ) 00HISTORY CURRICULUM SPECIALISTS: YES
002 CSPECH-N (2          ) 10HISTORY CURRICULUM SPECIALISTS: NO
003 CSPECH-? (M          ) 01HISTORY CURRICULUM SPECIALISTS: MISSING, DOES NOT APPLY

CONDITIONING VARIABLE ID:  TSUB0001
DESCRIPTION:                IS CLASS ASSIGNED BY ABILITY?
GRADES/ASSESSMENTS:         N04, N08
CONDITIONING VAR LABEL:     CLASSABL
NAEP ID:                    T051601                    TOTAL NUMBER OF SPECIFIED CONTRASTS:     3
TYPE OF CONTRAST:           CLASS                      NUMBER OF INDEPENDENT CONTRASTS:         2

001 CLASAB-Y (1          ) 00CLASS BY ABILITY: YES
002 CLASAB-N (2          ) 10CLASS BY ABILITY: NO
003 CLASAB-? (M          ) 01CLASS BY ABILITY: MISSING

CONDITIONING VARIABLE ID:  TSUB0002
DESCRIPTION:                WHAT IS THE ABILITY LEVEL OF THE STUDENTS IN THIS CLASS?
GRADES/ASSESSMENTS:         N04, N08
CONDITIONING VAR LABEL:     CLABLLEV
NAEP ID:                    T051701                    TOTAL NUMBER OF SPECIFIED CONTRASTS:     5
TYPE OF CONTRAST:           CLASS                      NUMBER OF INDEPENDENT CONTRASTS:         4

001 CLABLV-H (1          ) 0000                        CLASS ABILITY LEVEL: PRIMARILY HIGH
002 CLASIZ-A (2          ) 1000                        CLASS ABILITY LEVEL: PRIMARILY AVERAGE
003 CLASIZ-L (3          ) 0100                        CLASS ABILITY LEVEL: PRIMARILY LOW
004 CLASIZ-M (4          ) 0010                        CLASS ABILITY LEVEL: PRIMARILY MIXED
005 CLASIZ-? (M          ) 0001                        CLASS ABILITY LEVEL: MISSING

CONDITIONING VARIABLE ID:  TSUB0003
DESCRIPTION:                WHICH BEST DESCRIBES THE FOCUS OF YOUR SOCIAL STUDIES/HISTORY TEACHING?
GRADES/ASSESSMENTS:         N04
CONDITIONING VAR LABEL:     FOCUS
NAEP ID:                    T051801                    TOTAL NUMBER OF SPECIFIED CONTRASTS:     7
TYPE OF CONTRAST:           CLASS                      NUMBER OF INDEPENDENT CONTRASTS:         6

001 FOC-STAT (1          ) 000000                      FOCUS: STATE HISTORY
002 FOC-US-H (2          ) 100000                      FOCUS: U.S. HISTORY
003 FOC-SS   (3          ) 010000                      FOCUS: SOCIAL STUDIES
004 FOC-SS+  (4          ) 001000                      FOCUS: SOCIAL STUDIES INTEGRATED WITH OTHER
                                                       DISCIPLINES
005 FOC-GEOG (5          ) 000100                      FOCUS: GEOGRAPHY
006 FOC-OTH  (6          ) 000010                      FOCUS: OTHER
007 FOC-MISS (M          ) 000001                      FOCUS: MISSING

CONDITIONING VARIABLE ID:  TSUB0004
```

Table C-5 (continued)
Conditioning Variables Specific to the U.S. History Assessment

```
DESCRIPTION:                 ABOUT HOW MUCH TIME DO YOU SPEND WITH THIS CLASS ON UNITED STATES HISTOY
                             INSTRUCTION PER WEEK?
GRADES/ASSESSMENTS:          N04
CONDITIONING VAR LABEL:      AMT_TIME
NAEP ID:                     T051901                  TOTAL NUMBER OF SPECIFIED CONTRASTS:    5
TYPE OF CONTRAST:            CLASS                    NUMBER OF INDEPENDENT CONTRASTS:        4

001 TIME-30- (1            ) 0000                     HISTORY TIME: 30 MINUTES OR LESS
002 TIME-45M (2            ) 1000                     HISTORY TIME: 45 MINUTES
003 TIME-60M (3            ) 0100                     HISTORY TIME: 60 MINUTES
004 TIME-90+ (4            ) 0010                     HISTORY TIME: 90 MINUTES OR MORE
005 TIME-MIS (M            ) 0001                     HISTORY TIME: MISSING

CONDITIONING VARIABLE ID:    TSUB0005
DESCRIPTION:                 ABOUT HOW MANY WEEKS DURING THE SCHOOL YEAR DO YOU SPEND ON THE PRE-1815 PERIOD?
GRADES/ASSESSMENTS:          N04, N08
CONDITIONING VAR LABEL:      WKS1815
NAEP ID:                     T052001                  TOTAL NUMBER OF SPECIFIED CONTRASTS:    5
TYPE OF CONTRAST:            CLASS                    NUMBER OF INDEPENDENT CONTRASTS:        4

001 WK1815-1 (1            ) 0000                     WEEKS ON PRE-1815 PERIOD:  6 OR MORE WEEKS
002 WK1815-2 (2            ) 1000                     WEEKS ON PRE-1815 PERIOD:  3 TO 5 WEEKS
003 WK1815-3 (3            ) 0100                     WEEKS ON PRE-1815 PERIOD:  1-2 WEEKS
004 WK1815-4 (4            ) 0010                     WEEKS ON PRE-1815 PERIOD:  NOT COVERED IN
                                                      CURRICULUM
005 WK1815-? (M            ) 0001                     WEEKS ON PRE-1815 PERIOD:  MISSING

CONDITIONING VARIABLE ID:    TSUB0006
DESCRIPTION:                 ABOUT HOW MANY WEEKS DURING THE SCHOOL YEAR DO YOU SPEND ON THE 1815 TO 1865
                             PERIOD?
GRADES/ASSESSMENTS:          N04, N08
CONDITIONING VAR LABEL:      WKS1865
NAEP ID:                     T052002                  TOTAL NUMBER OF SPECIFIED CONTRASTS:    5
TYPE OF CONTRAST:            CLASS                    NUMBER OF INDEPENDENT CONTRASTS:        4

001 WK1865-1 (1            ) 0000                     WEEKS ON 1815 TO 1865 PERIOD: 6 OR MORE WEEKS
002 WK1865-2 (2            ) 1000                     WEEKS ON 1815 TO 1865 PERIOD: 3 TO 5 WEEKS
003 WK1865-3 (3            ) 0100                     WEEKS ON 1815 TO 1865 PERIOD: 1-2 WEEKS
004 WK1865-4 (4            ) 0010                     WEEKS ON 1815 TO 1865 PERIOD: NOT COVERED IN
                                                      CURRICULUM
005 WK1865-? (M            ) 0001                     WEEKS ON 1815 TO 1865 PERIOD: MISSING

CONDITIONING VARIABLE ID:    TSUB0007
DESCRIPTION:                 ABOUT HOW MANY WEEKS DURING THE SCHOOL YEAR DO YOU SPEND ON THE 1865 TO 1945
                             PERIOD?
GRADES/ASSESSMENTS:          N04, N08
CONDITIONING VAR LABEL:      WKS1945
NAEP ID:                     T052003                  TOTAL NUMBER OF SPECIFIED CONTRASTS:    5
TYPE OF CONTRAST:            CLASS                    NUMBER OF INDEPENDENT CONTRASTS:        4

001 WK1945-1 (1            ) 0000                     WEEKS ON 1865 TO 1945 PERIOD: 6 OR MORE WEEKS
002 WK1945-2 (2            ) 1000                     WEEKS ON 1865 TO 1945 PERIOD: 3 TO 5 WEEKS
003 WK1945-3 (3            ) 0100                     WEEKS ON 1865 TO 1945 PERIOD: 1-2 WEEKS
004 WK1945-4 (4            ) 0010                     WEEKS ON 1865 TO 1945 PERIOD: NOT COVERED IN
                                                      CURRICULUM
005 WK1945-? (M            ) 0001                     WEEKS ON 1865 TO 1945 PERIOD: MISSING

CONDITIONING VARIABLE ID:    TSUB0008
DESCRIPTION:                 ABOUT HOW MANY WEEKS DURING THE SCHOOL YEAR DO YOU SPEND ON THE 1945 TO THE PRESENT
                             PERIOD?
GRADES/ASSESSMENTS:          N04, N08
```

Table C-5 (continued)

Conditioning Variables Specific to the U.S. History Assessment

```
CONDITIONING VAR LABEL:     WKSCURR
NAEP ID:                    T052004                     TOTAL NUMBER OF SPECIFIED CONTRASTS:     5
TYPE OF CONTRAST:           CLASS                       NUMBER OF INDEPENDENT CONTRASTS:         4

001 WKCURR-1 (1          ) 0000                         WEEKS ON 1945 TO THE PRESENT PERIOD: 6 OR MORE
                                                        WEEKS
002 WKCURR-2 (2          ) 1000                         WEEKS ON 1945 TO THE PRESENT PERIOD: 3 TO 5 WEEKS
003 WKCURR-3 (3          ) 0100                         WEEKS ON 1945 TO THE PRESENT PERIOD: 1-2 WEEKS
004 WKCURR-4 (4          ) 0010                         WEEKS ON 1945 TO THE PRESENT PERIOD: NOT COVERED
                                                        IN CURRICULUM
005 WKCURR-? (M          ) 0001                         WEEKS ON 1945 TO THE PRESENT PERIOD: MISSING

CONDITIONING VARIABLE ID:   TSUB0009
DESCRIPTION:                ABOUT HOW MUCH DO YOU EMPHASIZE THE DEVELOPMENT OF AMERICAN DEMOCRACY AND POLITICS?
GRADES/ASSESSMENTS:         N04, N08
CONDITIONING VAR LABEL:     EMPDEMO
NAEP ID:                    T052101                     TOTAL NUMBER OF SPECIFIED CONTRASTS:     4
TYPE OF CONTRAST:           CLASS                       NUMBER OF INDEPENDENT CONTRASTS:         3

001 EMPDEM-H (1          ) 000                          EMPHASIZE DEMOCRACY: HEAVY
002 EMPDEM-M (2          ) 100                          EMPHASIZE DEMOCRACY: MODERATE
003 EMPDEM-L (3          ) 010                          EMPHASIZE DEMOCRACY: LITTLE OR NO
004 EMPDEM-? (M          ) 001                          EMPHASIZE DEMOCRACY: MISSING

CONDITIONING VARIABLE ID:   TSUB0010
DESCRIPTION:                ABOUT HOW MUCH DO YOU EMPHASIZE THE GATHERING AND INTERACTION OF PEOPLES,IDEAS AND
                            CULTURES?
GRADES/ASSESSMENTS:         N04, N08
CONDITIONING VAR LABEL:     EMPCULT
NAEP ID:                    T052102                     TOTAL NUMBER OF SPECIFIED CONTRASTS:     4
TYPE OF CONTRAST:           CLASS                       NUMBER OF INDEPENDENT CONTRASTS:         3

001 EMPCUL-H (1          ) 000                          EMPHASIZE CULTURE: HEAVY
002 EMPCUL-M (2          ) 100                          EMPHASIZE CULTURE: MODERATE
003 EMPCUL-L (3          ) 010                          EMPHASIZE CULTURE: LITTLE OR NO
004 EMPCUL-? (M          ) 001                          EMPHASIZE CULTURE: MISSING

CONDITIONING VARIABLE ID:   TSUB0011
DESCRIPTION:                ABOUT HOW MUCH DO YOU EMPHASIZE THE IMPACT OF ECONOMIC AND TECHNOLOGICAL CHANGES?

GRADES/ASSESSMENTS:         N04, N08
CONDITIONING VAR LABEL:     EMPTECH
NAEP ID:                    T052103                     TOTAL NUMBER OF SPECIFIED CONTRASTS:     4

TYPE OF CONTRAST:           CLASS                       NUMBER OF INDEPENDENT CONTRASTS:         3


001 EMPTEC-H (1          ) 000                          EMPHASIZE TECHNOLOGY: HEAVY
002 EMPTEC-M (2          ) 100                          EMPHASIZE TECHNOLOGY: MODERATE
003 EMPTEC-L (3          ) 010                          EMPHASIZE TECHNOLOGY: LITTLE OR NO
004 EMPTEC-? (M          ) 001                          EMPHASIZE TECHNOLOGY: MISSING

CONDITIONING VARIABLE ID:   TSUB0012
DESCRIPTION:                ABOUT HOW MUCH DO YOU EMPHASIZE THE CHANGING ROLE OF THE UNITED STATES IN THE
                            WORLD?
GRADES/ASSESSMENTS:         N04, N08
CONDITIONING VAR LABEL:     EMPWRLE
NAEP ID:                    T052104                     TOTAL NUMBER OF SPECIFIED CONTRASTS:     4
TYPE OF CONTRAST:           CLASS                       NUMBER OF INDEPENDENT CONTRASTS:         3

001 EMPWRL-H (1          ) 000                          EMPHASIZE CHANGING WORLD ROLE: HEAVY
002 EMPWRL-M (2          ) 100                          EMPHASIZE CHANGING WORLD ROLE: MODERATE
```

Table C-5 (continued)
Conditioning Variables Specific to the U.S. History Assessment

```
003 EMPWRL-L (3           ) 010                       EMPHASIZE CHANGING WORLD ROLE: LITTLE OR NO
004 EMPWRL-? (M           ) 001                       EMPHASIZE CHANGING WORLD ROLE: MISSING

CONDITIONING VARIABLE ID:  TSUB0013
DESCRIPTION:               ABOUT HOW MUCH U.S. HISTORY HOMEWORK DO YOU ASSIGN A STUDENT IN THIS CLASS EACH
                           WEEK?
GRADES/ASSESSMENTS:        N04, N08
CONDITIONING VAR LABEL:    HISTHWK
NAEP ID:                   T052201                    TOTAL NUMBER OF SPECIFIED CONTRASTS:     6
TYPE OF CONTRAST:          CLASS                      NUMBER OF INDEPENDENT CONTRASTS:         5

001 HISHWK-N (1           ) 00000                     HOW MUCH HISTORY HOMEWORK: NONE
002 HISHWK-H (2           ) 10000                     HOW MUCH HISTORY HOMEWORK: 1/2 HOUR
003 HISHWK-1 (3           ) 01000                     HOW MUCH HISTORY HOMEWORK: 1 HOUR
004 HISHWK-2 (4           ) 00100                     HOW MUCH HISTORY HOMEWORK: 2 HOURS
005 HISHW-2+ (5           ) 00010                     HOW MUCH HISTORY HOMEWORK: MORE THAN 2 HOURS
006 HISHWK-? (M           ) 00001                     HOW MUCH HISTORY HOMEWORK: MISSING

CONDITIONING VARIABLE ID:  TSUB0014
DESCRIPTION:               HOW OFTEN DO YOU HAVE YOUR CLASS READ MATERIAL FROM TEXTBOOKS WHEN TEACHING
                           HISTORY?
GRADES/ASSESSMENTS:        N04, N08
CONDITIONING VAR LABEL:    READTEXT
NAEP ID:                   T052301                    TOTAL NUMBER OF SPECIFIED CONTRASTS:     5
TYPE OF CONTRAST:          CLASS                      NUMBER OF INDEPENDENT CONTRASTS:         4

001 RDTXT--D (1           ) 0000                      HOW OFTEN READ TEXTBOOKS: ALMOST EVERY DAY
002 RDTXT--W (2           ) 1000                      HOW OFTEN READ TEXTBOOKS: ONCE OR TWICE A WEEK
003 RDTXT--M (3           ) 0100                      HOW OFTEN READ TEXTBOOKS: ONCE OR TWICE A MONTH
004 RDTXT--N (4           ) 0010                      HOW OFTEN READ TEXTBOOKS: NEVER OR HARDLY EVER
005 RDTXT--? (M           ) 0001                      HOW OFTEN READ TEXTBOOKS: MISSING

CONDITIONING VARIABLE ID:  TSUB0015
DESCRIPTION:               HOW OFTEN DO YOU HAVE YOUR CLASS READ MATERIALS OTHER THAN TEXTBOOKS WHEN TEACHING
                           HISTORY?
GRADES/ASSESSMENTS:        N04, N08
CONDITIONING VAR LABEL:    READOTH
NAEP ID:                   T052302                    TOTAL NUMBER OF SPECIFIED CONTRASTS:     5
TYPE OF CONTRAST:          CLASS                      NUMBER OF INDEPENDENT CONTRASTS:         4

001 RDOTH--D (1           ) 0000                      HOW OFTEN READ OTHER THAN TEXTBOOKS: ALMOST EVERY
                                                      DAY
002 RDOTH--W (2           ) 1000                      HOW OFTEN READ OTHER THAN TEXTBOOKS: ONCE OR TWICE
                                                      A WEEK
003 RDOTH--M (3           ) 0100                      HOW OFTEN READ OTHER THAN TEXTBOOKS: ONCE OR TWICE
                                                      A MONTH
004 RDOTH--N (4           ) 0010                      HOW OFTEN READ OTHER THAN TEXTBOOKS: NEVER OR
                                                      HARDLY EVER
005 RDOTH--? (M           ) 0001                      HOW OFTEN READ OTHER THAN TEXTBOOKS: MISSING

CONDITIONING VARIABLE ID:  TSUB0016
DESCRIPTION:               HOW OFTEN DO YOU HAVE YOUR CLASS READ PRIMARY HISTORICAL DOCUMENTS WHEN TEACHING
                           HISTORY?
GRADES/ASSESSMENTS:        N04, N08
CONDITIONING VAR LABEL:    USEPRIM
NAEP ID:                   T052303                    TOTAL NUMBER OF SPECIFIED CONTRASTS:     5
TYPE OF CONTRAST:          CLASS                      NUMBER OF INDEPENDENT CONTRASTS:         4
```

Table C-5 (continued)
Conditioning Variables Specific to the U.S. History Assessment

```
001 USEPRM-D (1          ) 0000                         HOW OFTEN READ PRIMARY HISTORICAL DOCUMENTS:
                                                        ALMOST EVERY DAY
002 USEPRM-W (2          ) 1000                         HOW OFTEN READ PRIMARY HISTORICAL DOCUMENTS: ONCE
                                                        OR TWICE A WEEK
003 USEPRM-M (3          ) 0100                         HOW OFTEN READ PRIMARY HISTORICAL DOCUMENTS: ONCE
                                                        OR TWICE A MONTH
004 USEPRM-N (4          ) 0010                         HOW OFTEN READ PRIMARY HISTORICAL DOCUMENTS: NEVER
                                                        OR HARDLY EVER
005 USEPRM-? (M          ) 0001                         HOW OFTEN READ PRIMARY HISTORICAL DOCUMENTS:
                                                        MISSING

CONDITIONING VARIABLE ID:  TSUB0017
DESCRIPTION:               HOW OFTEN DO YOU HAVE YOUR CLASS TAKE TESTS OR QUIZZES?
GRADES/ASSESSMENTS:        N04, N08
CONDITIONING VAR LABEL:    TESTS
NAEP ID:                   T052304                      TOTAL NUMBER OF SPECIFIED CONTRASTS:      5
TYPE OF CONTRAST:          CLASS                        NUMBER OF INDEPENDENT CONTRASTS:          4

001 TESTS--D (1          ) 0000                         HOW OFTEN TAKE TESTS OR QUIZZES: ALMOST EVERY DAY
002 TESTS--W (2          ) 1000                         HOW OFTEN TAKE TESTS OR QUIZZES: ONCE OR TWICE A
                                                        WEEK
003 TESTS--M (3          ) 0100                         HOW OFTEN TAKE TESTS OR QUIZZES: ONCE OR TWICE A
                                                        MONTH
004 TESTS--N (4          ) 0010                         HOW OFTEN TAKE TESTS OR QUIZZES: NEVER OR HARDLY
                                                        EVER
005 TESTS--? (M          ) 0001                         HOW OFTEN TAKE TESTS OR QUIZZES: MISSING

CONDITIONING VARIABLE ID:  TSUB0018
DESCRIPTION:               HOW OFTEN DO YOU HAVE YOUR CLASS DISCUSS THE MATERIAL STUDIED?
GRADES/ASSESSMENTS:        N04, N08
CONDITIONING VAR LABEL:    DISCUSS
NAEP ID:                   T052305                      TOTAL NUMBER OF SPECIFIED CONTRASTS:      5
TYPE OF CONTRAST:          CLASS                        NUMBER OF INDEPENDENT CONTRASTS:          4

001 DISCUS-D (1          ) 0000                         HOW OFTEN DISCUSS MATERIALS: ALMOST EVERY DAY
002 DISCUS-W (2          ) 1000                         HOW OFTEN DISCUSS MATERIALS: ONCE OR TWICE A WEEK
003 DISCUS-M (3          ) 0100                         HOW OFTEN DISCUSS MATERIALS: ONCE OR TWICE A MONTH
004 DISCUS-N (4          ) 0010                         HOW OFTEN DISCUSS MATERIALS: NEVER OR HARDLY EVER
005 DISCUS-? (M          ) 0001                         HOW OFTEN DISCUSS MATERIALS: MISSING

CONDITIONING VARIABLE ID:  TSUB0019
DESCRIPTION:               HOW OFTEN DO YOU HAVE YOUR CLASS WRITING SHORT ANSWERS TO QUESTIONS?
GRADES/ASSESSMENTS:        N04, N08
CONDITIONING VAR LABEL:    WRTSHORT
NAEP ID:                   T052306                      TOTAL NUMBER OF SPECIFIED CONTRASTS:      5
TYPE OF CONTRAST:          CLASS                        NUMBER OF INDEPENDENT CONTRASTS:          4

001 WRTSHT-D (1          ) 0000                         HOW OFTEN WRITING SHORT ANSWERS: ALMOST EVERY DAY
002 WRTSHT-W (2          ) 1000                         HOW OFTEN WRITING SHORT ANSWERS: ONCE OR TWICE A
                                                        WEEK
003 WRTSHT-M (3          ) 0100                         HOW OFTEN WRITING SHORT ANSWERS: ONCE OR TWICE A
                                                        MONTH
004 WRTSHT-N (4          ) 0010                         HOW OFTEN WRITING SHORT ANSWERS: NEVER OR HARDLY
                                                        EVER
005 WRTSHT-? (M          ) 0001                         HOW OFTEN WRITING SHORT ANSWERS: MISSING

CONDITIONING VARIABLE ID:  TSUB0020
DESCRIPTION:               HOW OFTEN DO YOU HAVE YOUR CLASS WRITING REPORTS?
GRADES/ASSESSMENTS:        N04, N08
CONDITIONING VAR LABEL:    WRTREPRT
NAEP ID:                   T052307                      TOTAL NUMBER OF SPECIFIED CONTRASTS:      5
```

Table C-5 (continued)

Conditioning Variables Specific to the U.S. History Assessment

```
TYPE OF CONTRAST:          CLASS                         NUMBER OF INDEPENDENT CONTRASTS:        4

001 WRTRPT-D (1          ) 0000                          HOW OFTEN WRITING REPORTS: ALMOST EVERY DAY
002 WRTRPT-W (2          ) 1000                          HOW OFTEN WRITING REPORTS: ONCE OR TWICE A WEEK
003 WRTRPT-M (3          ) 0100                          HOW OFTEN WRITING REPORTS: ONCE OR TWICE A MONTH
004 WRTRPT-N (4          ) 0010                          HOW OFTEN WRITING REPORTS: NEVER OR HARDLY EVER
005 WRTRPT-? (M          ) 0001                          HOW OFTEN WRITING REPORTS: MISSING

CONDITIONING VARIABLE ID:  TSUB0021
DESCRIPTION:               HOW OFTEN DO YOU HAVE YOUR CLASS WORKING IN GROUPS?
GRADES/ASSESSMENTS:        N04, N08
CONDITIONING VAR LABEL:    GRPPROJ
NAEP ID:                   T052308                       TOTAL NUMBER OF SPECIFIED CONTRASTS:    5
TYPE OF CONTRAST:          CLASS                         NUMBER OF INDEPENDENT CONTRASTS:        4

001 GRPPRJ-D (1          ) 0000                          HOW OFTEN WORKING IN GROUPS: ALMOST EVERY DAY
002 GRPPRJ-W (2          ) 1000                          HOW OFTEN WORKING IN GROUPS: ONCE OR TWICE A WEEK
003 GRPPRJ-M (3          ) 0100                          HOW OFTEN WORKING IN GROUPS: ONCE OR TWICE A MONTH
004 GRPPRJ-N (4          ) 0010                          HOW OFTEN WORKING IN GROUPS: NEVER OR HARDLY EVER
005 GRPPRJ-? (M          ) 0001                          HOW OFTEN WORKING IN GROUPS: MISSING

CONDITIONING VARIABLE ID:  TSUB0022
DESCRIPTION:               HOW OFTEN DO YOU HAVE YOUR CLASS GIVE ORAL REPORTS?
GRADES/ASSESSMENTS:        N04, N08
CONDITIONING VAR LABEL:    ORALREP
NAEP ID:                   T052309                       TOTAL NUMBER OF SPECIFIED CONTRASTS:    5
TYPE OF CONTRAST:          CLASS                         NUMBER OF INDEPENDENT CONTRASTS:        4

001 ORALRP-D (1          ) 0000                          HOW OFTEN GIVE ORAL REPORTS: ALMOST EVERY DAY
002 ORALRP-W (2          ) 1000                          HOW OFTEN GIVE ORAL REPORTS: ONCE OR TWICE A WEEK
003 ORALRP-M (3          ) 0100                          HOW OFTEN GIVE ORAL REPORTS: ONCE OR TWICE A MONTH
004 ORALRP-N (4          ) 0010                          HOW OFTEN GIVE ORAL REPORTS: NEVER OR HARDLY EVER
005 ORALRP-? (M          ) 0001                          HOW OFTEN GIVE ORAL REPORTS: MISSING

CONDITIONING VARIABLE ID:  TSUB0023
DESCRIPTION:               HOW OFTEN DO YOU USE MAPS OR GLOBES?
GRADES/ASSESSMENTS:        N04, N08
CONDITIONING VAR LABEL:    USEMAPS
NAEP ID:                   T052310                       TOTAL NUMBER OF SPECIFIED CONTRASTS:    5
TYPE OF CONTRAST:          CLASS                         NUMBER OF INDEPENDENT CONTRASTS:        4

001 USEMAP-D (1          ) 0000                          HOW OFTEN USE MAPS OR GLOBES: ALMOST EVERY DAY
002 USEMAP-W (2          ) 1000                          HOW OFTEN USE MAPS OR GLOBES: ONCE OR TWICE A WEEK
003 USEMAP-M (3          ) 0100                          HOW OFTEN USE MAPS OR GLOBES: ONCE OR TWICE A
                                                         MONTH
004 USEMAP-N (4          ) 0010                          HOW OFTEN USE MAPS OR GLOBES: NEVER OR HARDLY EVER
005 USEMAP-? (M          ) 0001                          HOW OFTEN USE MAPS OR GLOBES: MISSING

CONDITIONING VARIABLE ID:  TSUB0024
DESCRIPTION:               HOW OFTEN DO YOU HAVE YOUR CLASS WATCH VIDEOS, MOVIES, OR FILMSTRIPS?
GRADES/ASSESSMENTS:        N04, N08
CONDITIONING VAR LABEL:    VIDEOS
NAEP ID:                   T052311                       TOTAL NUMBER OF SPECIFIED CONTRASTS:    5
TYPE OF CONTRAST:          CLASS                         NUMBER OF INDEPENDENT CONTRASTS:        4

001 VIDEOS-D (1          ) 0000                          HOW OFTEN WATCH VIDEOS: ALMOST EVERY DAY
002 VIDEOS-W (2          ) 1000                          HOW OFTEN WATCH VIDEOS: ONCE OR TWICE A WEEK
003 VIDEOS-M (3          ) 0100                          HOW OFTEN WATCH VIDEOS: ONCE OR TWICE A MONTH
004 VIDEOS-N (4          ) 0010                          HOW OFTEN WATCH VIDEOS: NEVER OR HARDLY EVER
005 VIDEOS-? (M          ) 0001                          HOW OFTEN WATCH VIDEOS: MISSING

CONDITIONING VARIABLE ID:  TSUB0025
DESCRIPTION:               HOW OFTEN DO YOU HAVE YOUR CLASS USE A COMPUTER?
GRADES/ASSESSMENTS:        N04, N08
```

Table C-5 (continued)

Conditioning Variables Specific to the U.S. History Assessment

```
CONDITIONING VAR LABEL:  COMPUTER
NAEP ID:                 T052312                        TOTAL NUMBER OF SPECIFIED CONTRASTS:   5
TYPE OF CONTRAST:        CLASS                          NUMBER OF INDEPENDENT CONTRASTS:       4

001 COMPUT-D (1        ) 0000                           HOW OFTEN USE A COMPUTER: ALMOST EVERY DAY
002 COMPUT-W (2        ) 1000                           HOW OFTEN USE A COMPUTER: ONCE OR TWICE A WEEK
003 COMPUT-M (3        ) 0100                           HOW OFTEN USE A COMPUTER: ONCE OR TWICE A MONTH
004 COMPUT-N (4        ) 0010                           HOW OFTEN USE A COMPUTER: NEVER OR HARDLY EVER
005 COMPUT-? (M        ) 0001                           HOW OFTEN USE A COMPUTER: MISSING

CONDITIONING VARIABLE ID: TSUB0026
DESCRIPTION:             HOW OFTEN DO YOU HAVE YOUR CLASS GO ON FIELD TRIPS, OR HAVE OUTSIDE SPEAKERS?
GRADES/ASSESSMENTS:      N04, N08
CONDITIONING VAR LABEL:  TRIPS
NAEP ID:                 T052313                        TOTAL NUMBER OF SPECIFIED CONTRASTS:   5
TYPE OF CONTRAST:        CLASS                          NUMBER OF INDEPENDENT CONTRASTS:       4

001 TRIPS--D (1        ) 0000                           HOW OFTEN GO ON FIELD TRIPS: ALMOST EVERY DAY
002 TRIPS--W (2        ) 1000                           HOW OFTEN GO ON FIELD TRIPS: ONCE OR TWICE A WEEK
003 TRIPS--M (3        ) 0100                           HOW OFTEN GO ON FIELD TRIPS: ONCE OR TWICE A MONTH
004 TRIPS--N (4        ) 0010                           HOW OFTEN GO ON FIELD TRIPS: NEVER OR HARDLY EVER
005 TRIPS--? (M        ) 0001                           HOW OFTEN GO ON FIELD TRIPS: MISSING

CONDITIONING VARIABLE ID: TSUB0027
DESCRIPTION:             HOW OFTEN DO YOU HAVE YOUR CLASS WORKING IN THE LIBRARY?
GRADES/ASSESSMENTS:      N04, N08
CONDITIONING VAR LABEL:  WKLIBR
NAEP ID:                 T052314                        TOTAL NUMBER OF SPECIFIED CONTRASTS:   5
TYPE OF CONTRAST:        CLASS                          NUMBER OF INDEPENDENT CONTRASTS:       4

001 WKLIBR-D (1        ) 0000                           HOW OFTEN WORKING IN LIBRARY: ALMOST EVERY DAY
002 WKLIBR-W (2        ) 1000                           HOW OFTEN WORKING IN LIBRARY: ONCE OR TWICE A WEEK
003 WKLIBR-M (3        ) 0100                           HOW OFTEN WORKING IN LIBRARY: ONCE OR TWICE A
                                                        MONTH
004 WKLIBR-N (4        ) 0010                           HOW OFTEN WORKING IN LIBRARY: NEVER OR HARDLY EVER
005 WKLIBR-? (M        ) 0001                           HOW OFTEN WORKING IN LIBRARY: MISSING

CONDITIONING VARIABLE ID: TSUB0028
DESCRIPTION:             HOW OFTEN DO YOU SEND OR TAKE YOUR CLASS TO THE LIBRARY?
GRADES/ASSESSMENTS:      N04, N08
CONDITIONING VAR LABEL:  GOLIBR
NAEP ID:                 T052401                        TOTAL NUMBER OF SPECIFIED CONTRASTS:   6
TYPE OF CONTRAST:        CLASS                          NUMBER OF INDEPENDENT CONTRASTS:       5

001 GOLIB--D (1        ) 00000                          HOW OFTEN TAKE TO LIBRARY: ALMOST EVERY DAY
002 GOLIB--W (2        ) 10000                          HOW OFTEN TAKE TO LIBRARY: ONCE OR TWICE A WEEK
003 GOLIB--M (3        ) 01000                          HOW OFTEN TAKE TO LIBRARY: ONCE OR TWICE A MONTH
004 GOLIB--N (4        ) 00100                          HOW OFTEN TAKE TO LIBRARY: NEVER OR HARDLY EVER
005 GOLIB-NO (5        ) 00010                          HOW OFTEN TAKE TO LIBRARY: THERE IS NO LIBRAY AT
                                                        THIS SCHOOL
006 GOLIBR-? (M        ) 00001                          HOW OFTEN TAKE TO LIBRARY: MISSING

CONDITIONING VARIABLE ID: TSUB0029
DESCRIPTION:             HOW OFTEN DO YOU ASSIGN STUDENTS TO READ A BOOK FROM THE LIBRARY?
GRADES/ASSESSMENTS:      N04, N08
CONDITIONING VAR LABEL:  LIBBOOK
NAEP ID:                 T052402                        TOTAL NUMBER OF SPECIFIED CONTRASTS:   6
TYPE OF CONTRAST:        CLASS                          NUMBER OF INDEPENDENT CONTRASTS:       5

001 LIBOOK-D (1        ) 00000                          HOW OFTEN ASSIGN BOOK FROM LIBRARY: ALMOST EVERY
                                                        DAY
002 LIBOOK-W (2        ) 10000                          HOW OFTEN ASSIGN BOOK FROM LIBRARY: ONCE OR TWICE
                                                        A WEEK
```

Table C-5 (continued)

Conditioning Variables Specific to the U.S. History Assessment

```
003 LIBOOK-M (3          ) 01000                     HOW OFTEN ASSIGN BOOK FROM LIBRARY: ONCE OR TWICE
                                                     A MONTH
004 LIBOOK-N (4          ) 00100                     HOW OFTEN ASSIGN BOOK FROM LIBRARY: NEVER OR
                                                     HARDLY EVER
005 LIBOK-NO (5          ) 00010                     HOW OFTEN ASSIGN BOOK FROM LIBRARY: THERE IS NO
                                                     LIBRAY AT THIS SCHOOL
006 LIBOOK-? (M          ) 00001                     HOW OFTEN ASSIGN BOOK FROM LIBRARY: MISSING

CONDITIONING VARIABLE ID:  TSUB0030
DESCRIPTION:               WHICH BEST DESCRIBES AVAILABILITY OF COMPUTERS FOR USE BY STUDENTS  IN YOUR HISTORY
                           CLASS?
GRADES/ASSESSMENTS:        N04, N08
CONDITIONING VAR LABEL:    COMPAVA
NAEP ID:                   T052501                   TOTAL NUMBER OF SPECIFIED CONTRASTS:     4
TYPE OF CONTRAST:          CLASS                     NUMBER OF INDEPENDENT CONTRASTS:         3

001 COMPAV-N (1          ) 000                       CPMPUTERS NOT AVAILABLE
002 COMPAV-D (2          ) 100                       COMPUTERS AVAILABLE, BUT DIFFICULT TO USE.
003 COMPAV-C (3          ) 010                       COMPUTERS AVAILABLE IN CLASSROOM.
004 COMPAV-? (M          ) 001                       AVAILABILITY OF COMPUTERS ARE MISSING

CONDITIONING VARIABLE ID:  TSUB0031
DESCRIPTION:               HOW OFTEN DO YOU USE MULTIPLE CHOICE TESTS TO ASSESS STUDENT PROGRESS IN HISTORY?
GRADES/ASSESSMENTS:        N04, N08
CONDITIONING VAR LABEL:    MCTESTS
NAEP ID:                   T052601                   TOTAL NUMBER OF SPECIFIED CONTRASTS:     5
TYPE OF CONTRAST:          CLASS                     NUMBER OF INDEPENDENT CONTRASTS:         4

001 MCTEST-W (1          ) 0000                      USE MULTIPLE CHOICE TESTS:  ONCE OR TWICE A WEEK
002 MCTEST-M (2          ) 1000                      USE MULTIPLE CHOICE TESTS:  ONCE OR TWICE A MONTH
003 MCTEST-Y (3          ) 0100                      USE MULTIPLE CHOICE TESTS:  ONCE OR TWICE A YEAR
004 MCTEST-N (4          ) 0010                      USE MULTIPLE CHOICE TESTS:  NEVER OF HARDLY EVER
005 MCTEST-? (M          ) 0001                      USE MULTIPLE CHOICE TESTS:  MISSING

CONDITIONING VARIABLE ID:  TSUB0032
DESCRIPTION:               HOW OFTEN DO YOU USE SHORT OR LONG WRITTEN RESPONSE TO ASSESS STUDENT PROGRESS IN
                           HISTORY?
GRADES/ASSESSMENTS:        N04, N08
CONDITIONING VAR LABEL:    WRTRESP
NAEP ID:                   T052602                   TOTAL NUMBER OF SPECIFIED CONTRASTS:     5
TYPE OF CONTRAST:          CLASS                     NUMBER OF INDEPENDENT CONTRASTS:         4

001 WRTRSP-W (1          ) 0000                      USE SHORT OR LONG WRITTEN RESPONSES: ONCE OR TWICE
                                                     A WEEK
002 WRTRSP-M (2          ) 1000                      USE SHORT OR LONG WRITTEN RESPONSES: ONCE OR TWICE
                                                     A MONTH
003 WRTRSP-Y (3          ) 0100                      USE SHORT OR LONG WRITTEN RESPONSES: ONCE OR TWICE
                                                     A YEAR
004 WRTRSP-N (4          ) 0010                      USE SHORT OR LONG WRITTEN RESPONSES: NEVER OF
                                                     HARDLY EVER
005 WRTRSP-? (M          ) 0001                      USE SHORT OR LONG WRITTEN RESPONSES: MISSING

CONDITIONING VARIABLE ID:  TSUB0033
DESCRIPTION:               HOW OFTEN DO YOU USE INDIVIDUAL OR GROUP PROJECTS TO ASSESS STUDENT PROGRESS IN
                           HISTORY?
GRADES/ASSESSMENTS:        N04, N08
CONDITIONING VAR LABEL:    GRPPROJ
NAEP ID:                   T052603                   TOTAL NUMBER OF SPECIFIED CONTRASTS:     5
TYPE OF CONTRAST:          CLASS                     NUMBER OF INDEPENDENT CONTRASTS:         4

001 GRPPRJ-W (1          ) 0000                      USE PROJECTS: ONCE OR TWICE A WEEK
002 GRPPRJ-M (2          ) 1000                      USE PROJECTS: ONCE OR TWICE A MONTH
003 GRPPRJ-Y (3          ) 0100                      USE PROJECTS: ONCE OR TWICE A YEAR
004 GRPPRJ-N (4          ) 0010                      USE PROJECTS: NEVER OF HARDLY EVER
```

Table C-5 (continued)

Conditioning Variables Specific to the U.S. History Assessment

```
005 GRPPRJ-? (M          ) 0001                           USE PROJECTS: MISSING

CONDITIONING VARIABLE ID:  TSUB0034
DESCRIPTION:               HOW OFTEN DO YOU USE PORTFOLIO COLLECTIONS TO ASSESS STUDENT PROGRESS IN HISTORY?
GRADES/ASSESSMENTS:        N04, N08
CONDITIONING VAR LABEL:    TPORTF
NAEP ID:                   T052604                        TOTAL NUMBER OF SPECIFIED CONTRASTS:     5
TYPE OF CONTRAST:          CLASS                          NUMBER OF INDEPENDENT CONTRASTS:         4

001 TPORT--W (1          ) 0000                           USE PORTFOLIOS: ONCE OR TWICE A WEEK
002 TPORT--M (2          ) 1000                           USE PORTFOLIOS: ONCE OR TWICE A MONTH
003 TPORT--Y (3          ) 0100                           USE PORTFOLIOS: ONCE OR TWICE A YEAR
004 TPORT--N (4          ) 0010                           USE PORTFOLIOS: NEVER OF HARDLY EVER
005 TPORT--? (M          ) 0001                           USE PORTFOLIOS: MISSING

CONDITIONING VARIABLE ID:  TSUB0035
DESCRIPTION:               WHAT IS YOUR AVERAGE HISTORY CLASS SIZE?
GRADES/ASSESSMENTS:        N04
CONDITIONING VAR LABEL:    CLASSSZ4
NAEP ID:                   T052701                        TOTAL NUMBER OF SPECIFIED CONTRASTS:     6
TYPE OF CONTRAST:          CLASS                          NUMBER OF INDEPENDENT CONTRASTS:         5

001 CL401-20 (1          ) 00000                          CLASS SIZE:  1-20 STUDENTS   (FOURTH GRADE)
002 CL421-25 (2          ) 10000                          CLASS SIZE: 21-25 STUDENTS   (FOURTH GRADE)
003 CL426-30 (3          ) 01000                          CLASS SIZE: 26-30 STUDENTS   (FOURTH GRADE)
004 CL431-35 (4          ) 00100                          CLASS SIZE: 31-35 STUDENTS   (FOURTH GRADE)
005 CL436+   (5          ) 00010                          CLASS SIZE: 36 OR MORE STUDENTS  (FOURTH GRADE)
006 CL4-MISS (M          ) 00001                          CLASS SIZE: MISSING          (FOURTH GRADE)

CONDITIONING VARIABLE ID:  TSUB0036
DESCRIPTION:               WHAT IS THE PRIMARY EMPHASIS OF THIS COURSE?
GRADES/ASSESSMENTS:        N08
CONDITIONING VAR LABEL:    PEMPH
NAEP ID:                   T055101                        TOTAL NUMBER OF SPECIFIED CONTRASTS:     6
TYPE OF CONTRAST:          CLASS                          NUMBER OF INDEPENDENT CONTRASTS:         5

001 PEMP-HIS (1          ) 00000                          PRIMARY EMPHASIS: U.S.HISTORY
002 PEMP-GEO (2          ) 10000                          PRIMARY EMPHASIS: GEOGRAPHY
003 PEMP-SSO (3          ) 01000                          PRIMARY EMPHASIS: SOCIAL STUDIES
004 PEMP-SS+ (4          ) 00100                          PRIMARY EMPHASIS: SOCIAL STUDIES INTEGRATED WITH
                                                          OTHER DISCIPLINES
005 PEMP-OTH (5          ) 00010                          PRIMARY EMPHASIS: OTHER
006 PEMP-?   (M          ) 00001                          PRIMARY EMPHASIS: MISSING

CONDITIONING VARIABLE ID:  TSUB0037
DESCRIPTION:               WHAT PROPORTION OF THE COURSE YOU ARE TEACHING IS SPENT ON US HISTORY?
GRADES/ASSESSMENTS:        N08
CONDITIONING VAR LABEL:    %USHIST
NAEP ID:                   T055201                        TOTAL NUMBER OF SPECIFIED CONTRASTS:     5
TYPE OF CONTRAST:          CLASS                          NUMBER OF INDEPENDENT CONTRASTS:         4

001 HIST-<25 (1          ) 0000                           PROPORTION ON US HISTORY: LESS THAN 25%
002 HIST-<50 (2          ) 1000                           PROPORTION ON US HISTORY: 25% TO 50%
003 HIST-<75 (3          ) 0100                           PROPORTION ON US HISTORY: 50% TO 75%
004 HIST->75 (4          ) 0010                           PROPORTION ON US HISTORY: 75% OR MORE
005 HIST---? (M          ) 0001                           PROPORTION ON US HISTORY: MISSING

CONDITIONING VARIABLE ID:  TSUB0038
DESCRIPTION:               WHAT IS YOUR AVERAGE HISTORY CLASS SIZE?
GRADES/ASSESSMENTS:        N08
CONDITIONING VAR LABEL:    CLASSSZ8
NAEP ID:                   T055801                        TOTAL NUMBER OF SPECIFIED CONTRASTS:     6
TYPE OF CONTRAST:          CLASS                          NUMBER OF INDEPENDENT CONTRASTS:         5
```

Table C-5 (continued)

Conditioning Variables Specific to the U.S. History Assessment

```
001 CL801-20 (1          ) 00000                CLASS SIZE:  1-20 STUDENTS    (EIGHTH GRADE)
002 CL821-25 (2          ) 10000                CLASS SIZE: 21-25 STUDENTS    (EIGHTH GRADE)
003 CL826-30 (3          ) 01000                CLASS SIZE: 26-30 STUDENTS    (EIGHTH GRADE)
004 CL831-35 (4          ) 00100                CLASS SIZE: 31-35 STUDENTS    (EIGHTH GRADE)
005 CL836+   (5          ) 00010                CLASS SIZE: 36 OR MORE STUDENTS (EIGHTH GRADE)
006 CL8-MISS (M          ) 00001                CLASS SIZE: MISSING           (EIGHTH GRADE)

CONDITIONING VARIABLE ID:  TSUB0039
DESCRIPTION:               HOW WELL PREPARED DO YOU FEEL YOU ARE TO TEACH HISTORY AT THE ELEMENTARY-SCHOOL
                           LEVEL?
GRADES/ASSESSMENTS:        N04
CONDITIONING VAR LABEL:    PREPHIST
NAEP ID:                   T051201              TOTAL NUMBER OF SPECIFIED CONTRASTS:     5
TYPE OF CONTRAST:          CLASS                NUMBER OF INDEPENDENT CONTRASTS:         4

001 PRPHIS-V (1          ) 0000                 PREPARED TO TEACH HISTORY: VERY PREPARED
002 PRPHIS-A (2          ) 1000                 PREPARED TO TEACH HISTORY: ADEQUATELY PREPARED
003 PRPHIS-S (3          ) 0100                 PREPARED TO TEACH HISTORY: SOMEWHAT PREPARED
004 PRPHIS-U (4          ) 0010                 PREPARED TO TEACH HISTORY: UNPREPARED
005 PRPHIS-? (M          ) 0001                 PREPARED TO TEACH HISTORY: MISSING

CONDITIONING VARIABLE ID:  TSUB0040
DESCRIPTION:               HOW WELL PREPARED DO YOU FEEL YOU ARE TO TEACH SOCIAL STUDIES AT THE
                           ELEMENTARY-SCHOOL LEVEL?
GRADES/ASSESSMENTS:        N04
CONDITIONING VAR LABEL:    PREPSOCS
NAEP ID:                   T051202              TOTAL NUMBER OF SPECIFIED CONTRASTS:     5
TYPE OF CONTRAST:          CLASS                NUMBER OF INDEPENDENT CONTRASTS:         4

001 PRPSS--V (1          ) 0000                 PREPARED TO TEACH SOCIAL STUDIES: VERY PREPARED
002 PRPSS--A (2          ) 1000                 PREPARED TO TEACH SOCIAL STUDIES: ADEQUATELY
                                                PREPARED
003 PRPSS--S (3          ) 0100                 PREPARED TO TEACH SOCIAL STUDIES: SOMEWHAT
                                                PREPARED
004 PRPSS--U (4          ) 0010                 PREPARED TO TEACH SOCIAL STUDIES: UNPREPARED
005 PRPSS--? (M          ) 0001                 PREPARED TO TEACH SOCIAL STUDIES: MISSING

CONDITIONING VARIABLE ID:  TSUB0041
DESCRIPTION:               DURING THE PAST FIVE YEARS HAVE YOU TAKEN ANY UNIVERSITY COURSES IN HISTORY?
GRADES/ASSESSMENTS:        N04
CONDITIONING VAR LABEL:    UNIVCOUR
NAEP ID:                   T051301              TOTAL NUMBER OF SPECIFIED CONTRASTS:     2
TYPE OF CONTRAST:          CLASS                NUMBER OF INDEPENDENT CONTRASTS:         1

001 UNICOU-? (0,M        ) 0 HISTORY UNIVERSITY COURSES: NONE CHECKED
002 UNICOU-Y (1          ) 1 HISTORY UNIVERSITY COURSES: YES

CONDITIONING VARIABLE ID:  TSUB0042
DESCRIPTION:               DURING THE PAST FIVE YEARS HAVE YOU TAKEN ANY PROFESSIONAL DEVELOP'T WORKSHOPS IN
                           HISTORY?
GRADES/ASSESSMENTS:        N04
CONDITIONING VAR LABEL:    WORKSHOP
NAEP ID:                   T051302              TOTAL NUMBER OF SPECIFIED CONTRASTS:     2
TYPE OF CONTRAST:          CLASS                NUMBER OF INDEPENDENT CONTRASTS:         1

001 WRKSHP-? (0,M        ) 0 HISTORY PROFESSIONAL DEVELOPMENT WORKSHOPS OR SEMINARS: NONE CHECKED
002 WRKSHP-Y (1          ) 1 HISTORY PROFESSIONAL DEVELOPMENT WORKSHOPS OR SEMINARS: YES

CONDITIONING VARIABLE ID:  TSUB0043
DESCRIPTION:               DURING THE PAST FIVE YEARS HAVE YOU TAKEN ANY OTHER PROFESSIONAL ACTIVITIES IN
                           HISTORY?
GRADES/ASSESSMENTS:        N04
CONDITIONING VAR LABEL:    PROFACT
NAEP ID:                   T051303              TOTAL NUMBER OF SPECIFIED CONTRASTS:     2
```

Table C-5 (continued)

Conditioning Variables Specific to the U.S. History Assessment

```
TYPE OF CONTRAST:          CLASS                            NUMBER OF INDEPENDENT CONTRASTS:         1

001 PROACT-? (0,M        ) 0 HISTORY OTHER PROFESSIONAL ACTIVITIES: NONE CHECKED
002 PROACT-Y (1          ) 1 HISTORY OTHER PROFESSIONAL ACTIVITIES: YES

CONDITIONING VARIABLE ID:  TSUB0044
DESCRIPTION:               DURING THE PAST FIVE YEARS HAVE YOU TAKEN ANY OTHER PROFESSIONAL ACTIVITIES
                           HISTORY?
GRADES/ASSESSMENTS:        N04
CONDITIONING VAR LABEL:    NOED5
NAEP ID:                   T051304                          TOTAL NUMBER OF SPECIFIED CONTRASTS:     2
TYPE OF CONTRAST:          CLASS                            NUMBER OF INDEPENDENT CONTRASTS:         1

001 NOED5--? (0,M        ) 0 HISTORY OTHER PROFESSIONAL ACTIVITIES: NONE CHECKED
002 NOED5--Y (1          ) 1 HISTORY OTHER PROFESSIONAL ACTIVITIES: YES

CONDITIONING VARIABLE ID:  TSUB0045
DESCRIPTION:               DURING THE PAST FIVE YEARS HAVE YOU TAKEN ANY UNIVERSITY COURSES IN OTHER SOCIAL
                           STUDIES?
GRADES/ASSESSMENTS:        N04
CONDITIONING VAR LABEL:    UNIVCOUS
NAEP ID:                   T051305                          TOTAL NUMBER OF SPECIFIED CONTRASTS:     2
TYPE OF CONTRAST:          CLASS                            NUMBER OF INDEPENDENT CONTRASTS:         1

001 UNICOS-? (0,M        ) 0 OTHER SOCIAL STUDIES UNIVERSITY COURSES: NONE CHECKED
002 UNICOS-Y (1          ) 1 OTHER SOCIAL STUDIES UNIVERSITY COURSES: YES

CONDITIONING VARIABLE ID:  TSUB0046
DESCRIPTION:               DURING THE PAST FIVE YEARS HAVE YOU TAKEN ANY PROFESSIONAL DEVELOP'T WORKSHOPS IN
                           OTHER SOC STDS?
GRADES/ASSESSMENTS:        N04
CONDITIONING VAR LABEL:    WORKSHPS
NAEP ID:                   T051306                          TOTAL NUMBER OF SPECIFIED CONTRASTS:     2
TYPE OF CONTRAST:          CLASS                            NUMBER OF INDEPENDENT CONTRASTS:         1

001 WRKSHS-? (0,M        ) 0 OTHER SOCIAL STUDIES PROFESSIONAL DEVELOPMENT WORKSHOPS OR SEMINARS: NONE CHECKED
002 WRKSHS-Y (1          ) 1 OTHER SOCIAL STUDIES PROFESSIONAL DEVELOPMENT WORKSHOPS OR SEMINARS: YES

CONDITIONING VARIABLE ID:  TSUB0047
DESCRIPTION:               DURING THE PAST FIVE YEARS HAVE YOU TAKEN ANY OTHER PROFESSIONAL ACTIVITIES IN
                           OTHER SOC STUDIES?
GRADES/ASSESSMENTS:        N04
CONDITIONING VAR LABEL:    PROFCTS
NAEP ID:                   T051307                          TOTAL NUMBER OF SPECIFIED CONTRASTS:     2
TYPE OF CONTRAST:          CLASS                            NUMBER OF INDEPENDENT CONTRASTS:         1

001 PROACS-? (0,M        ) 0 OTHER SOCIAL STUDIES OTHER PROFESSIONAL ACTIVITIES: NONE CHECKED
002 PROACS-Y (1          ) 1 OTHER SOCIAL STUDIES OTHER PROFESSIONAL ACTIVITIES: YES

CONDITIONING VARIABLE ID:  TSUB0048
DESCRIPTION:               DURING THE PAST FIVE YEARS HAVE YOU TAKEN ANY OTHER PROFESSIONAL ACTIVITIES OTHER
                           SOCIAL STUDIES?
GRADES/ASSESSMENTS:        N04
CONDITIONING VAR LABEL:    NOED5SS
NAEP ID:                   T051308                          TOTAL NUMBER OF SPECIFIED CONTRASTS:     2
TYPE OF CONTRAST:          CLASS                            NUMBER OF INDEPENDENT CONTRASTS:         1

001 NOED5S-? (0,M        ) 0 OTHER SOCIAL STUDIES OTHER PROFESSIONAL ACTIVITIES: NONE CHECKED
002 NOED5S-Y (1          ) 1 OTHER SOCIAL STUDIES OTHER PROFESSIONAL ACTIVITIES: YES

CONDITIONING VARIABLE ID:  TSUB0049
DESCRIPTION:               WHICH OF THE FOLLOWING PLAYS THE CENTRAL ROLE IN YOUR TEACHING OF SOC STUD OR
                           HISTORY CURRICULUM?
```

Table C-5 (continued)

Conditioning Variables Specific to the U.S. History Assessment

```
GRADES/ASSESSMENTS:        N04
CONDITIONING VAR LABEL:    CRLECUR
NAEP ID:                   T051401                         TOTAL NUMBER OF SPECIFIED CONTRASTS:     7
TYPE OF CONTRAST:          CLASS                           NUMBER OF INDEPENDENT CONTRASTS:         6

001 CRLE-TXT (1          ) 000000                          CENTRAL ROLE IN CURRICULUM:  TEXTBOOK-ESTABLISHED
                                                           CURRICULUM GUIDES
002 CRLE-DST (2          ) 100000                          CENTRAL ROLE IN CURRICULUM:  DISTRICT-DEVELOPED
                                                           CURRICULUM GUIDES
003 CRLE-SCH (3          ) 010000                          CENTRAL ROLE IN CURRICULUM:  SCHOOL-DEVELOPED
                                                           CURRICULUM GUIDES
004 CRLE-IND (4          ) 001000                          CENTRAL ROLE IN CURRICULUM:  INDIVIDUALLY
                                                           DEVELOPED CURRICULUM GUIDES
005 CRLE-STA (5          ) 000100                          CENTRAL ROLE IN CURRICULUM:  STATE CURRICULUM
                                                           GUIDES OR FRAMEWORKS
006 CRLE-EXA (6          ) 000010                          CENTRAL ROLE IN CURRICULUM:  PUBLISHED EXAMS OR
                                                           ASSESSMENTS
007 CRLE-?   (M          ) 000001                          CENTRAL ROLE IN CURRICULUM:  MISSING

CONDITIONING VARIABLE ID:  TSUB0050
DESCRIPTION:               HOW MUCH DO YOU LIKE TEACHING UNITED STATES HISTORY?
GRADES/ASSESSMENTS:        N04
CONDITIONING VAR LABEL:    LKTCHHI
NAEP ID:                   T051501                         TOTAL NUMBER OF SPECIFIED CONTRASTS:     4
TYPE OF CONTRAST:          CLASS                           NUMBER OF INDEPENDENT CONTRASTS:         3

001 LKHIS-FV (1          ) 000                             US HISTORY: IS ONE OF MY FAVORITE SUBJECTS TO
                                                           TEACH.
002 LKHIS-OT (2          ) 100                             US HISTORY: I LIKE TEACHING OTHER SUBJECTS BETTER
003 LKHIS-NO (3          ) 010                             US HISTORY: I DO NOT ENJOY TEACHING UNITED STATES
                                                           HISTORY.
004 LKHIS-?  (M          ) 001                             US HISTORY: MISSING

CONDITIONING VARIABLE ID:  TSUB0051
DESCRIPTION:               HOW WELL PREPARED DO YOU FEEL YOU ARE TO TEACH UNITED STATES HISTORY AT THE JUNIOR
                           HIGH LEVEL?
GRADES/ASSESSMENTS:        N08
CONDITIONING VAR LABEL:    PRUSHIS
NAEP ID:                   T054701                         TOTAL NUMBER OF SPECIFIED CONTRASTS:     4
TYPE OF CONTRAST:          CLASS                           NUMBER OF INDEPENDENT CONTRASTS:         3

001 PRUSHI-W (1          ) 000                             PREPARED TO TEACH JR. HI HISTORY: WELL PREPARED
002 PRUSHI-S (2          ) 100                             PREPARED TO TEACH JR. HI HISTORY: SOMEWHAT
                                                           PREPARED
003 PRUSHI-U (3          ) 010                             PREPARED TO TEACH JR. HI HISTORY: UNPREPARED
004 PRUSHI-? (M          ) 001                             PREPARED TO TEACH JR. HI HISTORY: MISSING

CONDITIONING VARIABLE ID:  TSUB0052
DESCRIPTION:               DURING THE PAST FIVE YEARS HAVE YOU TAKEN ANY PROFESSIONAL DEVELOP'T WORKSHOPS IN
                           HISTORY  (8)?
GRADES/ASSESSMENTS:        N08
CONDITIONING VAR LABEL:    WORKSHP8
NAEP ID:                   T054801                         TOTAL NUMBER OF SPECIFIED CONTRASTS:     3
TYPE OF CONTRAST:          CLASS                           NUMBER OF INDEPENDENT CONTRASTS:         2

001 WRKSH8-Y (1          ) 00HISTORY PROFESSIONAL DEVELOPMENT WORKSHOPS OR SEMINARS: YES          (EIGHT
                           GRADE)
002 WRKSH8-N (2          ) 10HISTORY PROFESSIONAL DEVELOPMENT WORKSHOPS OR SEMINARS: NO           (EIGHT
                           GRADE)
003 WRKSH8-? (M          ) 01HISTORY PROFESSIONAL DEVELOPMENT WORKSHOPS OR SEMINARS: MISSING      (EIGHT
                           GRADE)

CONDITIONING VARIABLE ID:  TSUB0053
DESCRIPTION:               DURING THE PAST FIVE YEARS HAVE YOU TAKEN ANY UNIVERSITY COURSES IN HISTORY (8)?
```

Table C-5 (continued)

Conditioning Variables Specific to the U.S. History Assessment

```
GRADES/ASSESSMENTS:           N08
CONDITIONING VAR LABEL:       UNIVCOU8
NAEP ID:                      T054802                         TOTAL NUMBER OF SPECIFIED CONTRASTS:     3

TYPE OF CONTRAST:             CLASS                           NUMBER OF INDEPENDENT CONTRASTS:         2


001 UNIC08-Y (1          ) 00HISTORY UNIVERSITY COURSES: YES           (EIGHTH GRADE)
002 UNIC08-N (2          ) 10HISTORY UNIVERSITY COURSES: NO            (EIGHTH GRADE)
003 UNIC08-? (M          ) 01HISTORY UNIVERSITY COURSES: MISSING       (EIGHTH GRADE)

CONDITIONING VARIABLE ID:  TSUB0054
DESCRIPTION:               DURING THE PAST FIVE YEARS HAVE YOU TAKEN ANY OTHER PROFESSIONAL ACTIVITIES IN
                           HISTORY (8)?
GRADES/ASSESSMENTS:        N08
CONDITIONING VAR LABEL:    PROFAC8
NAEP ID:                   T054803                            TOTAL NUMBER OF SPECIFIED CONTRASTS:     3
TYPE OF CONTRAST:          CLASS                              NUMBER OF INDEPENDENT CONTRASTS:         2

001 PROAC8-Y (1          ) 00HISTORY OTHER PROFESSIONAL ACTIVITIES: NONE CHECKED   (EIGHTH GRADE)
002 PROAC8-N (2          ) 10HISTORY OTHER PROFESSIONAL ACTIVITIES: YES            (EIGHTH GRADE)
003 PROAC8-? (M          ) 01HISTORY OTHER PROFESSIONAL ACTIVITIES: YES            (EIGHTH GRADE)

CONDITIONING VARIABLE ID:  TSUB0055
DESCRIPTION:               WHICH OF THE FOLLOWING PLAYS THE CENTRAL ROLE IN YOUR TEACHING OF SOC STUD/HISTORY
                           CURRICULUM (8)?
GRADES/ASSESSMENTS:        N08
CONDITIONING VAR LABEL:    CRLECU8
NAEP ID:                   T054901                            TOTAL NUMBER OF SPECIFIED CONTRASTS:     7
TYPE OF CONTRAST:          CLASS                              NUMBER OF INDEPENDENT CONTRASTS:         6

001 CRL8-TXT (1          ) 000000                             CENTRAL ROLE IN CURRICULUM:  TEXTBOOK-ESTABLISHED
                                                              CURRICULUM GUIDES      (EIGHTH GRADE)
002 CRL8-DST (2          ) 100000                             CENTRAL ROLE IN CURRICULUM:  DISTRICT-DEVELOPED
                                                              CURRICULUM GUIDES         (EIGHTH GRADE)
003 CRL8-SCH (3          ) 010000                             CENTRAL ROLE IN CURRICULUM:  SCHOOL-DEVELOPED
                                                              CURRICULUM GUIDES            (EIGHTH GRADE)
004 CRL8-IND (4          ) 001000                             CENTRAL ROLE IN CURRICULUM:  INDIVIDUALLY
                                                              DEVELOPED CURRICULUM GUIDES    (EIGHTH GRADE)
005 CRL8-STA (5          ) 000100                             CENTRAL ROLE IN CURRICULUM:  STATE CURRICULUM
                                                              GUIDES OR FRAMEWORKS         (EIGHTH GRADE)
006 CRL8-EXA (6          ) 000010                             CENTRAL ROLE IN CURRICULUM:  PUBLISHED EXAMS OR
                                                              ASSESSMENTS               (EIGHTH GRADE)
007 CRL8-?   (M          ) 000001                             CENTRAL ROLE IN CURRICULUM:  MISSING     (EIGHTH
                                                              GRADE)

CONDITIONING VARIABLE ID:  TSUB0056
DESCRIPTION:               HOW MUCH DO YOU LIKE TEACHING UNITED STATES HISTORY (8)?
GRADES/ASSESSMENTS:        N08
CONDITIONING VAR LABEL:    LKTCHH8
NAEP ID:                   T055001                            TOTAL NUMBER OF SPECIFIED CONTRASTS:     5
TYPE OF CONTRAST:          CLASS                              NUMBER OF INDEPENDENT CONTRASTS:         4

001 LKHI8-FV (1          ) 0000                               US HISTORY: IS ONE OF MY FAVORITE SUBJECTS TO
                                                              TEACH.  (EIGHTH GRADE)
002 LKHI8-IN (2          ) 1000                               US HISTORY: I FIND IT INTERESTING TO TEACH.
                                                              (EIGHTH GRADE)
003 LKHI8-OT (3          ) 0100                               US HISTORY: I LIKE TEACHING OTHER SUBJECTS BETTER.
                                                              (EIGHTH GRADE)
004 LKHI8-NO (4          ) 0010                               US HISTORY: I DO NOT ENJOY TEACHING UNITED STATES
                                                              HISTORY. (EIGHTH GRADE)
005 LKHI8-?  (M          ) 0001                               US HISTORY: MISSING    (EIGHTH GRADE)
```

Table C-6

Conditioning Variables Specific to the Geography Assessment

```
CONDITIONING VARIABLE ID:  SUBJ0001
DESCRIPTION:               HOW OFTEN DO YOU USUALLY HAVE SOCIAL STUDIES OR HISTORY CLASSES IN SCHOOL?
GRADES/ASSESSMENTS:        N04
CONDITIONING VAR LABEL:    OFTEN
NAEP ID:                   G800401                    TOTAL NUMBER OF SPECIFIED CONTRASTS:     6
TYPE OF CONTRAST:          CLASS                      NUMBER OF INDEPENDENT CONTRASTS:         5

001 OFT_DAY  (1          ) 00000                      HOW OFTEN HISTORY:  EVERY DAY
002 OFT_3-4W (2          ) 10000                      HOW OFTEN HISTORY:  3-4 TIMES A WEEK
003 OFT_1-2W (3          ) 01000                      HOW OFTEN HISTORY:  1-2 TIMES A WEEK
004 OFT_<1W  (4          ) 00100                      HOW OFTEN HISTORY:  LESS THAN ONCE A WEEK
005 OFT_NEV  (5          ) 00010                      HOW OFTEN HISTORY:  NEVER, OR, HARDLY EVER
006 OFT_MIS  (M          ) 00001                      HOW OFTEN HISTORY:  MISSING

CONDITIONING VARIABLE ID:  SUBJ0002
DESCRIPTION:               DO EITHER YOU OR YOUR TEACHER SAVE YOUR GEOGRAPHY OR SOCIAL STUDIES WORK IN A
                           FOLDER OR A PORTFOLIO
GRADES/ASSESSMENTS:        N04, N08, N12
CONDITIONING VAR LABEL:    PORTFOLI
NAEP ID:                   G800501                    TOTAL NUMBER OF SPECIFIED CONTRASTS:     3
TYPE OF CONTRAST:          CLASS                      NUMBER OF INDEPENDENT CONTRASTS:         2

001 PORT-YES (1          ) 00  PORTFOLIO: YES
002 PORT--NO (2          ) 10  PORTFOLIO: NO
003 PORT-MIS (M          ) 01  PORTFOLIO: MISSING

CONDITIONING VARIABLE ID:  SUBJ0003
DESCRIPTION:               HOW OFTEN DO YOU USE COMPUTERS FOR STUDYING GEOGRAPHY?
GRADES/ASSESSMENTS:        N04, N08, N12
CONDITIONING VAR LABEL:    COMPGEOG
NAEP ID:                   G800601                    TOTAL NUMBER OF SPECIFIED CONTRASTS:     4
TYPE OF CONTRAST:          CLASS                      NUMBER OF INDEPENDENT CONTRASTS:         3

001 COMP-OFT (1          ) 000 USE COMPUTERS: OFTEN
002 COMP-SOM (2          ) 100 USE COMPUTERS: SOMETIMES
003 COMP--NO (3          ) 010 USE COMPUTERS: NEVER
004 COMP-MIS (M          ) 001 USE COMPUTERS: MISSING

CONDITIONING VARIABLE ID:  SUBJ0004
DESCRIPTION:               HOW MUCH GEOGRAPHY HOMEWORK DO YOU HAVE EACH WEEK?
GRADES/ASSESSMENTS:        N04, N08, N12
CONDITIONING VAR LABEL:    GEOGHWK
NAEP ID:                   G800701                    TOTAL NUMBER OF SPECIFIED CONTRASTS:     6
TYPE OF CONTRAST:          CLASS                      NUMBER OF INDEPENDENT CONTRASTS:         5

001 GEOHWK-0 (1          ) 00000                      GEOGRAPHY HOMEWORK: NONE
002 GEOHWK-H (2          ) 10000                      GEOGRAPHY HOMEWORK: 1/2 HOUR
003 GEOHWK-1 (3          ) 01000                      GEOGRAPHY HOMEWORK: 1 HOUR
004 GEOHWK-2 (4          ) 00100                      GEOGRAPHY HOMEWORK: 2 HOURS
005 GEOHWK>2 (5          ) 00010                      GEOGRAPHY HOMEWORK: > 2 HOURS
006 GEOHWK-? (M          ) 00001                      GEOGRAPHY HOMEWORK: MISSING

CONDITIONING VARIABLE ID:  SUBJ0005
DESCRIPTION:               HOW OFTEN DOES YOUR CLASS GO ON FIELD TRIPS THAT RELATE TO THE STUDY OF GEOGRAPHY?

GRADES/ASSESSMENTS:        N04, N08, N12
CONDITIONING VAR LABEL:    FLDTRPS
NAEP ID:                   G800801                    TOTAL NUMBER OF SPECIFIED CONTRASTS:     4
TYPE OF CONTRAST:          CLASS                      NUMBER OF INDEPENDENT CONTRASTS:         3

001 FLDTRP-O (1          ) 000 HOW OFTEN GO ON FIELD TRIPS: OFTEN
002 HISPEO-S (2          ) 100 HOW OFTEN GO ON FIELD TRIPS: SOMETIMES
003 HISPEO-N (3          ) 010 HOW OFTEN GO ON FIELD TRIPS: NEVER
004 HISPEO-? (M          ) 001 HOW OFTEN GO ON FIELD TRIPS: MISSING
```

Table C-6 (continued)

Conditioning Variables Specific to the Geography Assessment

```
CONDITIONING VARIABLE ID:  SUBJ0006
DESCRIPTION:               HOW OFTEN DO YOU COMPLETE PROJECTS RELATED TO STUDY OF GEOGRAPHY?
GRADES/ASSESSMENTS:        N04, N08, N12
CONDITIONING VAR LABEL:    GEOPRJS
NAEP ID:                   G800901                         TOTAL NUMBER OF SPECIFIED CONTRASTS:     4
TYPE OF CONTRAST:          CLASS                           NUMBER OF INDEPENDENT CONTRASTS:         3

001 GEOPRJ-O (1           ) 000 HOW OFTEN GO ON FIELD TRIPS: OFTEN
002 GEOPRJ-S (2           ) 100 HOW OFTEN GO ON FIELD TRIPS: SOMETIMES
003 GEOPRJ-N (3           ) 010 HOW OFTEN GO ON FIELD TRIPS: NEVER
004 GEOPRJ-? (M           ) 001 HOW OFTEN GO ON FIELD TRIPS: MISSING

CONDITIONING VARIAPLE ID:  SUBJ0007
DESCRIPTION:               HOW OFTEN DO YOU PARTICIPATE IN LOCAL, STATE, OR NATIONAL GEOGRAPHY CONTESTS?

GRADES/ASSESSMENTS:        N04, N08, N12
CONDITIONING VAR LABEL:    GEOCONT
NAEP ID:                   G801001                         TOTAL NUMBER OF SPECIFIED CONTRASTS:     4
TYPE OF CONTRAST:          CLASS                           NUMBER OF INDEPENDENT CONTRASTS:         3

001 GEOCNT-O (1           ) 000 HOW OFTEN PARTICIPATE IN CONTESTS: OFTEN
002 GEOCNT-S (2           ) 100 HOW OFTEN PARTICIPATE IN CONTESTS: SOMETIMES
003 GEOCNT-N (3           ) 010 HOW OFTEN PARTICIPATE IN CONTESTS: NEVER
004 GEOCNT-? (M           ) 001 HOW OFTEN PARTICIPATE IN CONTESTS: MISSING

CONDITIONING VARIABLE ID:  SUBJ0008
DESCRIPTION:               HOW OFTEN DO YOU USE FILMS, VIDEOS, OR FILMSTRIPS FOR STUDYING GEOGRAPHY?
GRADES/ASSESSMENTS:        N04, N08, N12
CONDITIONING VAR LABEL:    GEOVIDS
NAEP ID:                   G801101                         TOTAL NUMBER OF SPECIFIED CONTRASTS:     4
TYPE OF CONTRAST:          CLASS                           NUMBER OF INDEPENDENT CONTRASTS:         3

001 GEOVID-O (1           ) 000 HOW OFTEN USE VIDEOS: OFTEN
002 GEOVID-S (2           ) 100 HOW OFTEN USE VIDEOS: SOMETIMES
003 GEOVID-N (3           ) 010 HOW OFTEN USE VIDEOS: NEVER
004 GEOVID-? (M           ) 001 HOW OFTEN USE VIDEOS: MISSING

CONDITIONING VARIABLE ID:  SUBJ0009
DESCRIPTION:               HOW OFTEN DO YOU READ NEWSPAPERS, MAGAZINES, OR JOURNALS THAT RELATE TO GEOGRAPHY?

GRADES/ASSESSMENTS:        N04, N08, N12
CONDITIONING VAR LABEL:    READGEO
NAEP ID:                   G801201                         TOTAL NUMBER OF SPECIFIED CONTRASTS:     4

TYPE OF CONTRAST:          CLASS                           NUMBER OF INDEPENDENT CONTRASTS:         3


001 READGE-O (1           ) 000 HOW OFTEN READ NEWSPAPERS: OFTEN
002 READGE-S (2           ) 100 HOW OFTEN READ NEWSPAPERS: SOMETIMES
003 READGE-N (3           ) 010 HOW OFTEN READ NEWSPAPERS: NEVER
004 READGE-? (M           ) 001 HOW OFTEN READ NEWSPAPERS: MISSING

CONDITIONING VARIABLE ID:  SUBJ0010
DESCRIPTION:               HOW OFTEN DO YOU USE THE LIBRARY FOR STUDYING GEOGRAPHY?
GRADES/ASSESSMENTS:        N04, N08, N12
CONDITIONING VAR LABEL:    LIBRARY
NAEP ID:                   G801301                         TOTAL NUMBER OF SPECIFIED CONTRASTS:     4
TYPE OF CONTRAST:          CLASS                           NUMBER OF INDEPENDENT CONTRASTS:         3

001 LIBRAY-O (1           ) 000 HOW OFTEN USE LIBRARY: OFTEN
002 LIBRAY-S (2           ) 100 HOW OFTEN USE LIBRARY: SOMETIMES
003 LIBRAY-N (3           ) 010 HOW OFTEN USE LIBRARY: NEVER
004 LIBRAY-? (M           ) 001 HOW OFTEN USE LIBRARY: MISSING
```

Table C-6 (continued)

Conditioning Variables Specific to the Geography Assessment

```
CONDITIONING VARIABLE ID:  SUBJ0011
DESCRIPTION:               HOW MUCH DO YOU LIKE STUDYING GEOGRAPHY?
GRADES/ASSESSMENTS:        N04, N08, N12
CONDITIONING VAR LABEL:    LIKEGEO
NAEP ID:                   G801401                  TOTAL NUMBER OF SPECIFIED CONTRASTS:     4
TYPE OF CONTRAST:          CLASS                    NUMBER OF INDEPENDENT CONTRASTS:         3

001 LIKGEO-F (1          ) 000 HOW MUCH DO YOU LIKE STUDYING GEOGRAOHY: FAVORITE SUBJECT
002 LIKGEO-O (2          ) 100 HOW MUCH DO YOU LIKE STUDYING GEOGRAOHY: MOST OTHER SUBJECTS BETTER
003 LIKGEO-N (3          ) 010 HOW MUCH DO YOU LIKE STUDYING GEOGRAOHY: NEVER STUDIED GEOGRAPHY
004 LIKGEO-? (M          ) 001 HOW MUCH DO YOU LIKE STUDYING GEOGRAOHY: MISSING

CONDITIONING VARIABLE ID:  SUBJ0012
DESCRIPTION:               HOW OFTEN HAVE YOU STUDIED USING MAPS AND GLOBES?
GRADES/ASSESSMENTS:        N04, N08, N12
CONDITIONING VAR LABEL:    MAPSGLO
NAEP ID:                   G801501                  TOTAL NUMBER OF SPECIFIED CONTRASTS:     5
TYPE OF CONTRAST:          CLASS                    NUMBER OF INDEPENDENT CONTRASTS:         4

001 MAPSGL-D (1          ) 0000HOW OFTEN USE MAPS OR GLOBES: ALMOST EVERY DAY
002 MAPSGL-W (2          ) 1000HOW OFTEN USE MAPS OR GLOBES: ONCE OR TWICE A WEEK
003 MAPSGL-M (3          ) 0100HOW OFTEN USE MAPS OR GLOBES: ONCE OR TWICE A MONTH
004 MAPSGL-N (4          ) 0010HOW OFTEN USE MAPS OR GLOBES: NEVER OR HARDLY EVER
005 MAPSGL-? (M          ) 0001HOW OFTEN USE MAPS OR GLOBES: MISSING

CONDITIONING VARIABLE ID:  SUBJ0013
DESCRIPTION:               HOW OFTEN HAVE YOU MEASURED DISTANCES ON MAPS OR GLOBES?
GRADES/ASSESSMENTS:        N04, N08, N12
CONDITIONING VAR LABEL:    DISTANS
NAEP ID:                   G801502                  TOTAL NUMBER OF SPECIFIED CONTRASTS:     5
TYPE OF CONTRAST:          CLASS                    NUMBER OF INDEPENDENT CONTRASTS:         4

001 DISTNS-D (1          ) 0000HOW OFTEN MEASURE DISTANCES: ALMOST EVERY DAY
002 DISTNS-W (2          ) 1000HOW OFTEN MEASURE DISTANCES: ONCE OR TWICE A WEEK
003 DISTNS-M (3          ) 0100HOW OFTEN MEASURE DISTANCES: ONCE OR TWICE A MONTH
004 DISTNS-N (4          ) 0010HOW OFTEN MEASURE DISTANCES: NEVER OR HARDLY EVER
005 DISTNS-? (M          ) 0001HOW OFTEN MEASURE DISTANCES: MISSING

CONDITIONING VARIABLE ID:  SUBJ0014
DESCRIPTION:               HOW OFTEN FIND PLACES BY USING LATITUDE AND LONGITUDE?
GRADES/ASSESSMENTS:        N04, N08, N12
CONDITIONING VAR LABEL:    LATLONS
NAEP ID:                   G801503                  TOTAL NUMBER OF SPECIFIED CONTRASTS:     5
TYPE OF CONTRAST:          CLASS                    NUMBER OF INDEPENDENT CONTRASTS:         4

001 LATLGS-D (1          ) 0000HOW OFTEN USE LATITUDE AND LONGITUDE: ALMOST EVERY DAY
002 LATLGS-W (2          ) 1000HOW OFTEN USE LATITUDE AND LONGITUDE: ONCE OR TWICE A WEEK
003 LATLGS-M (3          ) 0100HOW OFTEN USE LATITUDE AND LONGITUDE: ONCE OR TWICE A MONTH
004 LATLGS-N (4          ) 0010HOW OFTEN USE LATITUDE AND LONGITUDE: NEVER OR HARDLY EVER
005 LATLGS-? (M          ) 0001HOW OFTEN USE LATITUDE AND LONGITUDE: MISSING

CONDITIONING VARIABLE ID:  SUBJ0015
DESCRIPTION:               HOW OFTEN STUDY CONTINENTS,OCEANS,RIVERS AND MOUNTAINS?
GRADES/ASSESSMENTS:        N04, N08, N12
CONDITIONING VAR LABEL:    CONTIN
NAEP ID:                   G801504                  TOTAL NUMBER OF SPECIFIED CONTRASTS:     5
TYPE OF CONTRAST:          CLASS                    NUMBER OF INDEPENDENT CONTRASTS:         4

001 CONTIN-D (1          ) 0000HOW OFTEN CONTINENTS, OCEANS, RIVERS AND MOUNTAINS: ALMOST EVERY DAY
002 CONTIN-W (2          ) 1000HOW OFTEN CONTINENTS, OCEANS, RIVERS AND MOUNTAINS: ONCE OR TWICE A WEEK
003 CONTIN-M (3          ) 0100HOW OFTEN CONTINENTS, OCEANS, RIVERS AND MOUNTAINS: ONCE OR TWICE A MONTH
004 CONTIN-N (4          ) 0010HOW OFTEN CONTINENTS, OCEANS, RIVERS AND MOUNTAINS: NEVER OR HARDLY EVER
```

Table C-6 (continued)

Conditioning Variables Specific to the Geography Assessment

```
005 CONTIN-? (M          ) 0001HOW OFTEN CONTINENTS, OCEANS, RIVERS AND MOUNTAINS: MISSING

CONDITIONING VARIABLE ID:  SUBJ0016
DESCRIPTION:               HOW OFTEN STUDY NATURAL RESOURCES?
GRADES/ASSESSMENTS:        N04, N08, N12
CONDITIONING VAR LABEL:    NATRES
NAEP ID:                   G801505                         TOTAL NUMBER OF SPECIFIED CONTRASTS:     5
TYPE OF CONTRAST:          CLASS                           NUMBER OF INDEPENDENT CONTRASTS:         4

001 NATRES-D (1          ) 0000HOW OFTEN NATURAL RESOURCES: ALMOST EVERY DAY
002 NATRES-W (2          ) 1000HOW OFTEN NATURAL RESOURCES: ONCE OR TWICE A WEEK
003 NATRES-M (3          ) 0100HOW OFTEN NATURAL RESOURCES: ONCE OR TWICE A MONTH
004 NATRES-N (4          ) 0010HOW OFTEN NATURAL RESOURCES: NEVER OR HARDLY EVER
005 NATRES-? (M          ) 0001HOW OFTEN NATURAL RESOURCES: MISSING

CONDITIONING VARIABLE ID:  SUBJ0017
DESCRIPTION:               HOW OFTEN STUDY WEATHER AND CLIMATE?
GRADES/ASSESSMENTS:        N04, N08, N12
CONDITIONING VAR LABEL:    WEATHER
NAEP ID:                   G801506                         TOTAL NUMBER OF SPECIFIED CONTRASTS:     5
TYPE OF CONTRAST:          CLASS                           NUMBER OF INDEPENDENT CONTRASTS:         4

001 WEATH--D (1          ) 0000HOW OFTEN WEATHER AND CLIMATE: ALMOST EVERY DAY
002 WEATH--W (2          ) 1000HOW OFTEN WEATHER AND CLIMATE: ONCE OR TWICE A WEEK
003 WEATH--M (3          ) 0100HOW OFTEN WEATHER AND CLIMATE: ONCE OR TWICE A MONTH
004 WEATH--N (4          ) 0010HOW OFTEN WEATHER AND CLIMATE: NEVER OR HARDLY EVER
005 WEATH--? (M          ) 0001HOW OFTEN WEATHER AND CLIMATE: MISSING

CONDITIONING VARIABLE ID:  SUBJ0018
DESCRIPTION:               HOW OFTEN STUDY COUNTRIES AND CULTURES?
GRADES/ASSESSMENTS:        N04, N08, N12
CONDITIONING VAR LABEL:    COUNTRY
NAEP ID:                   G801507                         TOTAL NUMBER OF SPECIFIED CONTRASTS:     5
TYPE OF CONTRAST:          CLASS                           NUMBER OF INDEPENDENT CONTRASTS:         4

001 CNTRY--D (1          ) 0000HOW OFTEN COUNTRIES AND CULTURES: ALMOST EVERY DAY
002 CNTRY--W (2          ) 1000HOW OFTEN COUNTRIES AND CULTURES: ONCE OR TWICE A WEEK
003 CNTRY--M (3          ) 0100HOW OFTEN COUNTRIES AND CULTURES: ONCE OR TWICE A MONTH
004 CNTRY--N (4          ) 0010HOW OFTEN COUNTRIES AND CULTURES: NEVER OR HARDLY EVER
005 CNTRY--? (M          ) 0001HOW OFTEN COUNTRIES AND CULTURES: MISSING

CONDITIONING VARIABLE ID:  SUBJ0019
DESCRIPTION:               HOW OFTEN STUDIED POPULATION GROWTH AND MOVEMENT?
GRADES/ASSESSMENTS:        N04, N08, N12
CONDITIONING VAR LABEL:    POPULAT
NAEP ID:                   G801508                         TOTAL NUMBER OF SPECIFIED CONTRASTS:     5
TYPE OF CONTRAST:          CLASS                           NUMBER OF INDEPENDENT CONTRASTS:         4

001 POPUL--D (1          ) 0000HOW OFTEN POPULATION GROWTH AND MOVEMENT: ALMOST EVERY DAY
002 POPUL--W (2          ) 1000HOW OFTEN POPULATION GROWTH AND MOVEMENT: ONCE OR TWICE A WEEK
003 POPUL--M (3          ) 0100HOW OFTEN POPULATION GROWTH AND MOVEMENT: ONCE OR TWICE A MONTH
004 POPUL--N (4          ) 0010HOW OFTEN POPULATION GROWTH AND MOVEMENT: NEVER OR HARDLY EVER
005 POPUL--? (M          ) 0001HOW OFTEN POPULATION GROWTH AND MOVEMENT: MISSING

CONDITIONING VARIABLE ID:  SUBJ0020
DESCRIPTION:               HOW OFTEN STUDIED CITIES, TOWNS, AND COMMUNITIES?
GRADES/ASSESSMENTS:        N04, N08, N12
CONDITIONING VAR LABEL:    CITIES
NAEP ID:                   G801509                         TOTAL NUMBER OF SPECIFIED CONTRASTS:     5
TYPE OF CONTRAST:          CLASS                           NUMBER OF INDEPENDENT CONTRASTS:         4

001 CITY---D (1          ) 0000HOW OFTEN CITIES, TOWNS AND COMMUNITIES: ALMOST EVERY DAY
002 CITY---W (2          ) 1000HOW OFTEN CITIES, TOWNS AND COMMUNITIES: ONCE OR TWICE A WEEK
003 CITY---M (3          ) 0100HOW OFTEN CITIES, TOWNS AND COMMUNITIES: ONCE OR TWICE A MONTH
```

BEST COPY AVAILABLE

Table C-6 (continued)

Conditioning Variables Specific to the Geography Assessment

```
004 CITY---N (4          ) 0010HOW OFTEN CITIES, TOWNS AND COMMUNITIES: NEVER OR HARDLY EVER
005 CITY---? (M          ) 0001HOW OFTEN CITIES, TOWNS AND COMMUNITIES: MISSING

CONDITIONING VARIABLE ID:  SUBJ0021
DESCRIPTION:               HOW OFTEN STUDIED ENVIRONMENTAL ISSUES?
GRADES/ASSESSMENTS:        N04, N08, N12
CONDITIONING VAR LABEL:    ENVIRN
NAEP ID:                   G801510                TOTAL NUMBER OF SPECIFIED CONTRASTS:    5
TYPE OF CONTRAST:          CLASS                  NUMBER OF INDEPENDENT CONTRASTS:        4

001 ENVIRN-D (1          ) 0000HOW OFTEN ENVIRONMENTAL ISSUES: ALMOST EVERY DAY
002 ENVIRN-W (2          ) 1000HOW OFTEN ENVIRONMENTAL ISSUES: ONCE OR TWICE A WEEK
003 ENVIRN-M (3          ) 0100HOW OFTEN ENVIRONMENTAL ISSUES: ONCE OR TWICE A MONTH
004 ENVIRN-N (4          ) 0010HOW OFTEN ENVIRONMENTAL ISSUES: NEVER OR HARDLY EVER
005 ENVIRN-? (M          ) 0001HOW OFTEN ENVIRONMENTAL ISSUES: MISSING

CONDITIONING VARIABLE ID:  SUBJ0022
DESCRIPTION:               HOW OFTEN STUDIED REGION AND REGIONAL CONCEPTS?
GRADES/ASSESSMENTS:        N04, N08, N12
CONDITIONING VAR LABEL:    REGIONG
NAEP ID:                   G801511                TOTAL NUMBER OF SPECIFIED CONTRASTS:    5
TYPE OF CONTRAST:          CLASS                  NUMBER OF INDEPENDENT CONTRASTS:        4

001 REGION-D (1          ) 0000HOW OFTEN REGION AND REGIONAL CONCEPTS: ALMOST EVERY DAY
002 REGION-W (2          ) 1000HOW OFTEN REGION AND REGIONAL CONCEPTS: ONCE OR TWICE A WEEK
003 REGION-M (3          ) 0100HOW OFTEN REGION AND REGIONAL CONCEPTS: ONCE OR TWICE A MONTH
004 REGION-N (4          ) 0010HOW OFTEN REGION AND REGIONAL CONCEPTS: NEVER OR HARDLY EVER
005 REGION-? (M          ) 0001HOW OFTEN REGION AND REGIONAL CONCEPTS: MISSING

CONDITIONING VARIABLE ID:  SUBJ0023
DESCRIPTION:               HOW OFTEN STUDIED DIFFUSION?
GRADES/ASSESSMENTS:        N12
CONDITIONING VAR LABEL:    DIFFUS
NAEP ID:                   G801512                TOTAL NUMBER OF SPECIFIED CONTRASTS:    5
TYPE OF CONTRAST:          CLASS                  NUMBER OF INDEPENDENT CONTRASTS:        4

001 DIFFUS-D (1          ) 0000HOW OFTEN STUDIED DIFFUSION: ALMOST EVERY DAY
002 DIFFUS-W (2          ) 1000HOW OFTEN STUDIED DIFFUSION: ONCE OR TWICE A WEEK
003 DIFFUS-M (3          ) 0100HOW OFTEN STUDIED DIFFUSION: ONCE OR TWICE A MONTH
004 DIFFUS-N (4          ) 0010HOW OFTEN STUDIED DIFFUSION: NEVER OR HARDLY EVER
005 DIFFUS-? (M          ) 0001HOW OFTEN STUDIED DIFFUSION: MISSING

CONDITIONING VARIABLE ID:  SUBJ0024
DESCRIPTION:               DID YOU TAKE, OR DO YOU EXPECT TO TAKE A GEOGRAPHY COURSE IN THE SIXTH GRADE?
GRADES/ASSESSMENTS:        N08
CONDITIONING VAR LABEL:    GEOGR6
NAEP ID:                   G801601                TOTAL NUMBER OF SPECIFIED CONTRASTS:    4
TYPE OF CONTRAST:          CLASS                  NUMBER OF INDEPENDENT CONTRASTS:        3

001 GEOGR6-Y (1          ) 000 GEOGRAPHY COURSE IN SIXTH GRADE: YES
002 GEOGR6-N (2          ) 100 GEOGRAPHY COURSE IN SIXTH GRADE: NO
003 GEOGR6-K (3          ) 010 GEOGRAPHY COURSE IN SIXTH GRADE: I DON'T KNOW
004 GEOGR6-? (M          ) 001 GEOGRAPHY COURSE IN SIXTH GRADE:  MISSING

CONDITIONING VARIABLE ID:  SUBJ0025
DESCRIPTION:               DID YOU TAKE, OR DO YOU EXPECT TO TAKE A GEOGRAPHY COURSE IN THE SEVENTH GRADE?

GRADES/ASSESSMENTS:        N08
CONDITIONING VAR LABEL:    GEOGR7
NAEP ID:                   G801602                TOTAL NUMBER OF SPECIFIED CONTRASTS:    4
TYPE OF CONTRAST:          CLASS                  NUMBER OF INDEPENDENT CONTRASTS:        3

001 GEOGR7-Y (1          ) 000 GEOGRAPHY COURSE IN SEVENTH GRADE: YES
002 GEOGR7-N (2          ) 100 GEOGRAPHY COURSE IN SEVENTH GRADE: NO
```

Table C-6 (continued)

Conditioning Variables Specific to the Geography Assessment

```
003 GEOGR7-K (3          ) 010 GEOGRAPHY COURSE IN SEVENTH GRADE: I DON'T KNOW
004 GEOGR7-? (M          ) 001 GEOGRAPHY COURSE IN SEVENTH GRADE: MISSING

CONDITIONING VARIABLE ID:  SUBJ0026
DESCRIPTION:               DID YOU TAKE, OR DO YOU EXPECT TO TAKE A GEOGRAPHY COURSE IN THE EIGHTH GRADE?
GRADES/ASSESSMENTS:        N08
CONDITIONING VAR LABEL:    GEOGR8
NAEP ID:                   G801603                    TOTAL NUMBER OF SPECIFIED CONTRASTS:     4
TYPE OF CONTRAST:          CLASS                      NUMBER OF INDEPENDENT CONTRASTS:         3

001 GEOGR8-Y (1          ) 000 GEOGRAPHY COURSE IN EIGHTH GRADE: YES
002 GEOGR8-N (2          ) 100 GEOGRAPHY COURSE IN EIGHTH GRADE: NO
003 GEOGR8-K (3          ) 010 GEOGRAPHY COURSE IN EIGHTH GRADE: I DON'T KNOW
004 GEOGR8-? (M          ) 001 GEOGRAPHY COURSE IN EIGHTH GRADE: MISSING

CONDITIONING VARIABLE ID:  SUBJ0027
DESCRIPTION:               DID YOU TAKE, OR DO YOU EXPECT TO TAKE A GEOGRAPHY COURSE IN THE NINTH GRADE?
GRADES/ASSESSMENTS:        N12
CONDITIONING VAR LABEL:    GEOGR9
NAEP ID:                   G801604                    TOTAL NUMBER OF SPECIFIED CONTRASTS:     4
TYPE OF CONTRAST:          CLASS                      NUMBER OF INDEPENDENT CONTRASTS:         3

001 GEOGR9-Y (1          ) 000 GEOGRAPHY COURSE IN NINTH GRADE: YES
002 GEOGR9-N (2          ) 100 GEOGRAPHY COURSE IN NINTH GRADE: NO
003 GEOGR9-K (3          ) 010 GEOGRAPHY COURSE IN NINTH GRADE: I DON'T KNOW
004 GEOGR9-? (M          ) 001 GEOGRAPHY COURSE IN NINTH GRADE: MISSING

CONDITIONING VARIABLE ID:  SUBJ0028
DESCRIPTION:               DID YOU TAKE, OR DO YOU EXPECT TO TAKE A GEOGRAPHY COURSE IN THE TENTH GRADE?

GRADES/ASSESSMENTS:        N12
CONDITIONING VAR LABEL:    GEOG10
NAEP ID:                   G801605                    TOTAL NUMBER OF SPECIFIED CONTRASTS:     4

TYPE OF CONTRAST:          CLASS                      NUMBER OF INDEPENDENT CONTRASTS:         3


001 GEOG10-Y (1          ) 000 GEOGRAPHY COURSE IN TENTH GRADE: YES
002 GEOG10-N (2          ) 100 GEOGRAPHY COURSE IN TENTH GRADE: NO
003 GEOG10-K (3          ) 010 GEOGRAPHY COURSE IN TENTH GRADE: I DON'T KNOW
004 GEOG10-? (M          ) 001 GEOGRAPHY COURSE IN TENTH GRADE: MISSING

CONDITIONING VARIABLE ID:  SUBJ0029
DESCRIPTION:               DID YOU TAKE, OR DO YOU EXPECT TO TAKE A GEOGRAPHY COURSE IN THE ELEVENTH GRADE?

GRADES/ASSESSMENTS:        N12
CONDITIONING VAR LABEL:    GEOG11
NAEP ID:                   G801606                    TOTAL NUMBER OF SPECIFIED CONTRASTS:     4
TYPE OF CONTRAST:          CLASS                      NUMBER OF INDEPENDENT CONTRASTS:         3

001 GEOG11-Y (1          ) 000 GEOGRAPHY COURSE IN ELEVENTH GRADE: YES
002 GEOG11-N (2          ) 100 GEOGRAPHY COURSE IN ELEVENTH GRADE: NO
003 GEOG11-K (3          ) 010 GEOGRAPHY COURSE IN ELEVENTH GRADE: I DON'T KNOW
004 GEOG11-? (M          ) 001 GEOGRAPHY COURSE IN ELEVENTH GRADE: MISSING

CONDITIONING VARIABLE ID:  SUBJ0030
DESCRIPTION:               DID YOU TAKE, OR DO YOU EXPECT TO TAKE A GEOGRAPHY COURSE IN THE TWELVETH GRADE?
GRADES/ASSESSMENTS:        N12
CONDITIONING VAR LABEL:    GEOG12
NAEP ID:                   G801607                    TOTAL NUMBER OF SPECIFIED CONTRASTS:     4
TYPE OF CONTRAST:          CLASS                      NUMBER OF INDEPENDENT CONTRASTS:         3
```

Table C-6 (continued)

Conditioning Variables Specific to the Geography Assessment

```
001 GEOG12-Y (1          ) 000 GEOGRAPHY COURSE IN TWELVETH GRADE: YES
002 GEOG12-N (2          ) 100 GEOGRAPHY COURSE IN TWELVETH GRADE: NO
003 GEOG12-K (3          ) 010 GEOGRAPHY COURSE IN TWELVETH GRADE: I DON'T KNOW
004 GEOG12-? (M          ) 001 GEOGRAPHY COURSE IN TWELVETH GRADE: MISSING

CONDITIONING VARIABLE ID:  SUBJ0031
DESCRIPTION:               HAVE YOU TAKEN AN EARTH SCIENCE COURSE?
GRADES/ASSESSMENTS:        N08
CONDITIONING VAR LABEL:    EARSCI
NAEP ID:                   G801701                TOTAL NUMBER OF SPECIFIED CONTRASTS:    3
TYPE OF CONTRAST:          CLASS                  NUMBER OF INDEPENDENT CONTRASTS:        2

001 EARSCI-Y (1          ) 00  EARTH SCIENCE COURSE: YES
002 EARSCI-N (2          ) 10  EARTH SCIENCE COURSE: NO
003 EARSCI-? (M          ) 01  EARTH SCIENCE COURSE: MISSING

CONDITIONING VARIABLE ID:  SUBJ0032
DESCRIPTION:               HAVE YOU TAKEN AN WORLD GEOGRAPHY COURSE?
GRADES/ASSESSMENTS:        N12
CONDITIONING VAR LABEL:    WLDGEO
NAEP ID:                   G801801                TOTAL NUMBER OF SPECIFIED CONTRASTS:    3
TYPE OF CONTRAST:          CLASS                  NUMBER OF INDEPENDENT CONTRASTS:        2

001 WLDGEO-Y (1          ) 00  WORLD GEOGRAPHY COURSE: YES
002 WLDGEO-N (2          ) 10  WORLD GEOGRAPHY COURSE: NO
003 WLDGEO-? (M          ) 01  WORLD GEOGRAPHY COURSE: MISSING

CONDITIONING VARIABLE ID:  SUBJ0033
DESCRIPTION:               HAVE YOU TAKEN AN U.S. GEOGRAPHY COURSE?
GRADES/ASSESSMENTS:        N12
CONDITIONING VAR LABEL:    USAGEO
NAEP ID:                   G801901                TOTAL NUMBER OF SPECIFIED CONTRASTS:    3
TYPE OF CONTRAST:          CLASS                  NUMBER OF INDEPENDENT CONTRASTS:        2

001 USAGEO-Y (1          ) 00  U.S. GEOGRAPHY COURSE: YES
002 USAGEO-N (2          ) 10  U.S. GEOGRAPHY COURSE: NO
003 USAGEO-? (M          ) 01  U.S. GEOGRAPHY COURSE: MISSING

CONDITIONING VARIABLE ID:  SUBJ0034
DESCRIPTION:               HAVE YOU TAKEN AN ADVANCED PLACEMENT COURSE IN HISTORY?
GRADES/ASSESSMENTS:        N12
CONDITIONING VAR LABEL:    ADVPLM
NAEP ID:                   G802001                TOTAL NUMBER OF SPECIFIED CONTRASTS:    3
TYPE OF CONTRAST:          CLASS                  NUMBER OF INDEPENDENT CONTRASTS:        2

001 ADVPLM-Y (1          ) 00  ADVANCED PLACEMENT COURSE IN HISTORY: YES
002 ADVPLM-N (2          ) 10  ADVANCED PLACEMENT COURSE IN HISTORY: NO
003 ADVPLM-? (M          ) 01  ADVANCED PLACEMENT COURSE IN HISTORY: MISSING

CONDITIONING VARIABLE ID:  SUBJ0035
DESCRIPTION:               ABOUT HOW MANY QUESTIONS DID YOU GET RIGHT ON THE GEOGRAPHY ASSESSMENT?
GRADES/ASSESSMENTS:        N04, N08, N12
CONDITIONING VAR LABEL:    #QUESTN+
NAEP ID:                   GM00101                TOTAL NUMBER OF SPECIFIED CONTRASTS:    5
TYPE OF CONTRAST:          CLASS                  NUMBER OF INDEPENDENT CONTRASTS:        4

001 #QUEST+1 (1          ) 0000NUMBER QUESTIONS RIGHT:  ALMOST ALL
002 #QUEST+2 (2          ) 1000NUMBER QUESTIONS RIGHT:  MORE THAN HALF
003 #QUEST+3 (3          ) 0100NUMBER QUESTIONS RIGHT:  ABOUT HALF
004 #QUEST+4 (4          ) 0010NUMBER QUESTIONS RIGHT:  LESS THAN HALF
005 #QUEST+? (M          ) 0001NUMBER QUESTIONS RIGHT:  MISSING
```

Table C-6 (continued)

Conditioning Variables Specific to the Geography Assessment

```
CONDITIONING VARIABLE ID:  SUBJ0036
DESCRIPTION:               HOW HARD WAS THIS GEOGRAPHY TEST COMPARED TO OTHERS?
GRADES/ASSESSMENTS:        N04, N08, N12
CONDITIONING VAR LABEL:    TEST DIF
NAEP ID:                   GM00201                    TOTAL NUMBER OF SPECIFIED CONTRASTS:    5
TYPE OF CONTRAST:          CLASS                      NUMBER OF INDEPENDENT CONTRASTS:        4

001 TESTDIF1 (1          ) 0000TEST DIFFICULTY:  MUCH HARDER THAN OTHERS
002 TESTDIF2 (2          ) 1000TEST DIFFICULTY:  HARDER THAN OTHERS
003 TESTDIF3 (3          ) 0100TEST DIFFICULTY:  ABOUT AS HARD AS OTHERS
004 TESTDIF4 (4          ) 0010TEST DIFFICULTY:  EASIER THAN OTHERS
005 TESTDIF? (M          ) 0001TEST DIFFICULTY:  MISSING

CONDITIONING VARIABLE ID:  SUBJ0037
DESCRIPTION:               HOW HARD DID YOU TRY ON THIS TEST COMPARED TO OTHER GEOGRAPHY TESTS?
GRADES/ASSESSMENTS:        N04, N08, N12
CONDITIONING VAR LABEL:    TEST EFF
NAEP ID:                   GM00301                    TOTAL NUMBER OF SPECIFIED CONTRASTS:    5
TYPE OF CONTRAST:          CLASS                      NUMBER OF INDEPENDENT CONTRASTS:        4

001 TESTEFF1 (1          ) 0000TEST EFFORT:  MUCH HARDER THAN OTHERS
002 TESTEFF2 (2          ) 1000TEST EFFORT:  HARDER THAN OTHERS
003 TESTEFF3 (3          ) 0100TEST EFFORT:  ABOUT AS HARD AS OTHERS
004 TESTEFF4 (4          ) 0010TEST EFFORT:  NOT AS HARD AS OTHERS
005 TESTEFF? (M          ) 0001TEST EFFORT:  MISSING

CONDITIONING VARIABLE ID:  SUBJ0038
DESCRIPTION:               HOW IMPORTANT WAS IT TO YOU TO DO WELL ON THE GEOGRAPHY TEST?
GRADES/ASSESSMENTS:        N04, N08, N12
CONDITIONING VAR LABEL:    TEST IMP
NAEP ID:                   GM00401                    TOTAL NUMBER OF SPECIFIED CONTRASTS:    5
TYPE OF CONTRAST:          CLASS                      NUMBER OF INDEPENDENT CONTRASTS:        4

001 TESTIMP1 (1          ) 0000TEST IMPORTANCE:  VERY IMPORTANT
002 TESTIMP2 (2          ) 1000TEST IMPORTANCE:  IMPORTANT
003 TESTIMP3 (3          ) 0100TEST IMPORTANCE:  SOMEWHAT IMPORTANT
004 TESTIMP4 (4          ) 0010TEST IMPORTANCE:  NOT VERY IMPORTANT
005 TESTIMP? (M          ) 0001TEST IMPORTANCE:  MISSING

CONDITIONING VARIABLE ID:  SUBJ0039
DESCRIPTION:               HOW OFTEN WERE YOU ASKED TO WRITE LONG ANSWERS ON GEOGRAPHY TESTS?
GRADES/ASSESSMENTS:        N04, N08, N12
CONDITIONING VAR LABEL:    LONG ANS
NAEP ID:                   GM00501                    TOTAL NUMBER OF SPECIFIED CONTRASTS:    5
TYPE OF CONTRAST:          CLASS                      NUMBER OF INDEPENDENT CONTRASTS:        4

001 LNGANSW1 (1          ) 0000LONG ANSWERS:  AT LEAST ONCE A WEEK
002 LNGANSW2 (2          ) 1000LONG ANSWERS:  ONCE OR TWICE A MONTH
003 LNGANSW3 (3          ) 0100LONG ANSWERS:  ONCE OR TWICE A YEAR
004 LNGANSW4 (4          ) 0010LONG ANSWERS:  NEVER
005 LNGANSW? (M          ) 0001LONG ANSWERS:  MISSING

CONDITIONING VARIABLE ID:  TCHR0001
DESCRIPTION:               TEACHER MATCH STATUS WITH STUDENT
GRADES/ASSESSMENTS:        N04, N08
CONDITIONING VAR LABEL:    T_MATCH
NAEP ID:                   TCHMTCH                    TOTAL NUMBER OF SPECIFIED CONTRASTS:    3
TYPE OF CONTRAST:          CLASS                      NUMBER OF INDEPENDENT CONTRASTS:        2

001 TMCH-NO  (1,M        ) 00   TEACHER MATCH:  NO MATCH
002 TMCH-PAR (2          ) 10   TEACHER MATCH:  PARTIAL MATCH
003 TMCH-COM (3          ) 01   TEACHER MATCH:  COMPLETE MATCH
```

BEST COPY AVAILABLE

Table C-6 (continued)

Conditioning Variables Specific to the Geography Assessment

```
CONDITIONING VARIABLE ID:  TCHR0002
DESCRIPTION:               TEACHER GENDER
GRADES/ASSESSMENTS:        N04, N08
CONDITIONING VAR LABEL:    T_GENDER
NAEP ID:                   T040001                    TOTAL NUMBER OF SPECIFIED CONTRASTS:     3
TYPE OF CONTRAST:          CLASS                      NUMBER OF INDEPENDENT CONTRASTS:         2

001 T_MALE   (1          ) 00  TEACHER GENDER:  MALE
002 T_FEMALE (2          ) 10  TEACHER GENDER:  FEMALE
003 T_SEX-?  (M          ) 01  TEACHER GENDER:  MISSING, DOES NOT APPLY

CONDITIONING VARIABLE ID:  TCHR0003
DESCRIPTION:               TEACHER RACE/ETHNICITY
GRADES/ASSESSMENTS:        N04, N08
CONDITIONING VAR LABEL:    T_RACE
NAEP ID:                   T050801                    TOTAL NUMBER OF SPECIFIED CONTRASTS:     7
TYPE OF CONTRAST:          CLASS                      NUMBER OF INDEPENDENT CONTRASTS:         6

001 T_WHITE  (1          ) 000000                     TEACHER RACE/ETHNICITY:  WHITE
002 T_BLACK  (2          ) 100000                     TEACHER RACE/ETHNICITY:  BLACK
003 T_HISP   (3          ) 010000                     TEACHER RACE/ETHNICITY:  HISPANIC
004 T_ASIAN  (4          ) 001000                     TEACHER RACE/ETHNICITY:  ASIAN
005 T_PAC IS (5          ) 000100                     TEACHER RACE/ETHNICITY:  PACIFIC ISLANDER
006 T_AM IND (6          ) 000010                     TEACHER RACE/ETHNICITY:  AMERICAN INDIAN/ALASKAN
                                                      NATIV
007 T_RACE-? (M          ) 000001                     TEACHER RACE/ETHNICITY:  MISSING, DOES NOT APPLY

CONDITIONING VARIABLE ID:  TCHR0004
DESCRIPTION:               YEARS TEACHING ELEMENTARY/SECONDARY SCHOOL
GRADES/ASSESSMENTS:        N04, N08
CONDITIONING VAR LABEL:    T_YRSEXP
NAEP ID:                   T040301                    TOTAL NUMBER OF SPECIFIED CONTRASTS:     6
TYPE OF CONTRAST:          CLASS                      NUMBER OF INDEPENDENT CONTRASTS:         5

001 T_YREXP1 (1          ) 00000                      YEARS TEACHING:  2 OR LESS YEARS
002 T_YREXP2 (2          ) 10000                      YEARS TEACHING:  3-5 YEARS
003 T_YREXP3 (3          ) 01000                      YEARS TEACHING:  6-10 YEARS
004 T_YREXP4 (4          ) 00100                      YEARS TEACHING:  11-24 YEARS
005 T_YREXP5 (5          ) 00010                      YEARS TEACHING:  25 OR MORE YEARS
006 T_YREXP? (M          ) 00001                      YEARS TEACHING:  MISSING, DOES NOT APPLY

CONDITIONING VARIABLE ID:  TCHR0005
DESCRIPTION:               TEACHER GENERAL CERTIFICATION (ELEMENTARY, MIDDLE/JUNIOR, HIGH SCHOOL EDUCATION)
GRADES/ASSESSMENTS:        N04, N08
CONDITIONING VAR LABEL:    CERT GEN
NAEP ID:                   T040501                    TOTAL NUMBER OF SPECIFIED CONTRASTS:     4
TYPE OF CONTRAST:          CLASS                      NUMBER OF INDEPENDENT CONTRASTS:         3

001 CERTG-Y  (1          ) 000 TEACHER GENERAL CERTIFICATION:  YES
002 CERTG-N  (2          ) 100 TEACHER GENERAL CERTIFICATION:  NO
003 CERTG-NS (3          ) 010 TEACHER GENERAL CERTIFICATION:  NOT OFFERED IN STATE
004 CERTG-?  (M          ) 001 TEACHER GENERAL CERTIFICATION:  MISSING, DOES NOT APPLY

CONDITIONING VARIABLE ID:  TCHR0006
DESCRIPTION:               TEACHER CERTIFICATION IN READING
GRADES/ASSESSMENTS:        N04
CONDITIONING VAR LABEL:    CERT RED
NAEP ID:                   T040502                    TOTAL NUMBER OF SPECIFIED CONTRASTS:     4
TYPE OF CONTRAST:          CLASS                      NUMBER OF INDEPENDENT CONTRASTS:         3
```

Table C-6 (continued)

Conditioning Variables Specific to the Geography Assessment

```
001 CERTR-Y  (1          ) 000 TEACHER READING CERTIFICATION:  YES
002 CERTR-N  (2          ) 100 TEACHER READING CERTIFICATION:  NO
003 CERTR-NS (3          ) 010 TEACHER READING CERTIFICATION:  NOT OFFERED IN STATE
004 CERTR-?  (M          ) 001 TEACHER READING CERTIFICATION:  MISSING, DOES NOT APPLY

CONDITIONING VARIABLE ID:  TCHR0007
DESCRIPTION:               TEACHER CERTIFICATION MIDDLE/JUNIOR HIGH SCHOOL/SECONDARY ENGLISH/LANGUAGE ARTS
GRADES/ASSESSMENTS:        N04
CONDITIONING VAR LABEL:    CERT LAN
NAEP ID:                   T040508                         TOTAL NUMBER OF SPECIFIED CONTRASTS:     4
TYPE OF CONTRAST:          CLASS                           NUMBER OF INDEPENDENT CONTRASTS:         3

001 CERTL-Y  (1          ) 000 TEACHER ENGLISH/LANGUAGE ARTS CERTIFICATION: YES
002 CERTL-N  (2          ) 100 TEACHER ENGLISH/LANGUAGE ARTS CERTIFICATION: NO
003 CERTL-NS (3          ) 010 TEACHER ENGLISH/LANGUAGE ARTS CERTIFICATION: NOT OFFERED IN STATE
004 CERTL-?  (M          ) 001 TEACHER ENGLISH/LANGUAGE ARTS CERTIFICATION: MISSING, DOES NOT APPLY

CONDITIONING VARIABLE ID:  TCHR0008
DESCRIPTION:               YEARS TEACHING SOCIAL STUDIES, HISTORY, OR GEOGRAPHY
GRADES/ASSESSMENTS:        N08
CONDITIONING VAR LABEL:    T_YRSEXP
NAEP ID:                   T054601                         TOTAL NUMBER OF SPECIFIED CONTRASTS:     6
TYPE OF CONTRAST:          CLASS                           NUMBER OF INDEPENDENT CONTRASTS:         5

001 T_YREXP1 (1          ) 00000                           YEARS TEACHING S.S., HIST, OR GEOG:  2 OR LESS
                                                           YEARS
002 T_YREXP2 (2          ) 10000                           YEARS TEACHING S.S., HIST, OR GEOG:  3-5 YEARS
003 T_YREXP3 (3          ) 01000                           YEARS TEACHING S.S., HIST, OR GEOG:  6-10 YEARS
004 T_YREXP4 (4          ) 00100                           YEARS TEACHING S.S., HIST, OR GEOG:  11-24 YEARS
005 T_YREXP5 (5          ) 00010                           YEARS TEACHING S.S., HIST, OR GEOG:  25 OR MORE
                                                           YEARS
006 T_YREXP? (M          ) 00001                           YEARS TEACHING S.S., HIST, OR GEOG:  MISSING, DOES
                                                           NOT APPLY

CONDITIONING VARIABLE ID:  TCHR0009
DESCRIPTION:               HOW WELL DOES YOUR SCHOOL PROVIDE YOU WITH INSTRUCTIONAL MATERIAL/RESOURCES YOU
                           NEED?
GRADES/ASSESSMENTS:        N04
CONDITIONING VAR LABEL:    RESOURCE
NAEP ID:                   T041201                         TOTAL NUMBER OF SPECIFIED CONTRASTS:     5
TYPE OF CONTRAST:          CLASS                           NUMBER OF INDEPENDENT CONTRASTS:         4

001 RESOURC1 (1          ) 0000RESOURCES:  GET ALL
002 RESOURC2 (2          ) 1000RESOURCES:  GET MOST
003 RESOURC3 (3          ) 0100RESOURCES:  GET SOME
004 RESOURC4 (4          ) 0010RESOURCES:  DON'T GET
005 RESOURC? (M          ) 0001RESOURCES:  MISSING, DOES NOT APPLY

CONDITIONING VARIABLE ID:  TCHR0010
DESCRIPTION:               TEACHER UNDERGRADUATE MAJOR IN EDUCATION
GRADES/ASSESSMENTS:        N04, N08
CONDITIONING VAR LABEL:    UGRAD ED
NAEP ID:                   T040701                         TOTAL NUMBER OF SPECIFIED CONTRASTS:     2
TYPE OF CONTRAST:          CLASS                           NUMBER OF INDEPENDENT CONTRASTS:         1

001 UGR ED-? (0,M        ) 0   TEACHER UNDERGRAD EDUCATION MAJOR:  MISSING, DOES NOT APPLY
002 UGR ED-Y (1          ) 1   TEACHER UNDERGRAD EDUCATION MAJOR:  YES

CONDITIONING VARIABLE ID:  TCHR0011
DESCRIPTION:               TEACHER GRADUATE MAJOR IN EDUCATION
GRADES/ASSESSMENTS:        N04, N08
CONDITIONING VAR LABEL:    GRAD ED
```

Table C-6 (continued)

Conditioning Variables Specific to the Geography Assessment

```
NAEP ID:                    T040801                     TOTAL NUMBER OF SPECIFIED CONTRASTS:     2
TYPE OF CONTRAST:           CLASS                       NUMBER OF INDEPENDENT CONTRASTS:         1

001 GRA ED-? (0,M        ) 0   TEACHER GRADUATE EDUCATION MAJOR:  MISSING, DOES NOT APPLY
002 GRA ED-Y (1          ) 1   TEACHER GRADUATE EDUCATION MAJOR:  YES

CONDITIONING VARIABLE ID:  TCHR0012
DESCRIPTION:                NO TEACHER GRADUATE-LEVEL STUDY
GRADES/ASSESSMENTS:         N04, N08
CONDITIONING VAR LABEL:     NO GRAD
NAEP ID:                    T040806                     TOTAL NUMBER OF SPECIFIED CONTRASTS:     2
TYPE OF CONTRAST:           CLASS                       NUMBER OF INDEPENDENT CONTRASTS:         1

001 NOGRAD-? (0,M        ) 0   NO TEACHER GRADUATE STUDY:  MISSING, DOES NOT APPLY
002 NOGRAD-Y (1          ) 1   NO TEACHER GRADUATE STUDY:  YES

CONDITIONING VARIABLE ID:  TCHR0013
DESCRIPTION:                HOW MANY YEARS IN TOTAL HAVE YOU TAUGHT READING? (4TH GRADE)
GRADES/ASSESSMENTS:         N04
CONDITIONING VAR LABEL:     T4REDYRS
NAEP ID:                    T049901                     TOTAL NUMBER OF SPECIFIED CONTRASTS:     6
TYPE OF CONTRAST:           CLASS                       NUMBER OF INDEPENDENT CONTRASTS:         5

001 T4REDYR1 (1          ) 00000                        YEARS TEACHING READING:  2 YEARS OR LESS
002 T4REDYR2 (2          ) 10000                        YEARS TEACHING READING:  3-5 YEARS
003 T4REDYR3 (3          ) 01000                        YEARS TEACHING READING:  6-10 YEARS
004 T4REDYR4 (4          ) 00100                        YEARS TEACHING READING:  11-24 YEARS
005 T4REDYR5 (5          ) 00010                        YEARS TEACHING READING:  25 YEARS OR MORE
006 T4REDYR? (M          ) 00001                        YEARS TEACHING READING:  MISSING, DOES NOT APPLY

CONDITIONING VARIABLE ID:  TCHR0014
DESCRIPTION:                DURING THE LAST YEAR IN SOCIAL STUDIES STAFF DEVELOPEMENT WORKSHOPS OR SEMINARS?
GRADES/ASSESSMENTS:         N04, N08
CONDITIONING VAR LABEL:     TCHWRKSH
NAEP ID:                    T050301                     TOTAL NUMBER OF SPECIFIED CONTRASTS:     6
TYPE OF CONTRAST:           CLASS                       NUMBER OF INDEPENDENT CONTRASTS:         5

001 TWKSP-0  (1          ) 00000                        TEACHER WORKSHOPS OR SEMINARS:  NONE
002 TWKSP-<6 (2          ) 10000                        TEACHER WORKSHOPS OR SEMINARS:  LESS THAN 6 HOURS
003 TWKSP-15 (3          ) 01000                        TEACHER WORKSHOPS OR SEMINARS:  6 - 15 HOURS
004 TWKSP-35 (4          ) 00100                        TEACHER WORKSHOPS OR SEMINARS:  16 - 35 HOURS
005 TWKSP>35 (5          ) 00010                        TEACHER WORKSHOPS OR SEMINARS:  MORE THAN 35 HOURS
006 TWKSP-?  (M          ) 00001                        TEACHER WORKSHOPS OR SEMINARS:  MISSING, DOES NOT
                                                        APPLY

CONDITIONING VARIABLE ID:  TCHR0015
DESCRIPTION:                WHAT TYPE OF TEACHING CERTIFICATION DO YOU HAVE THAT IS RECOGNIZED BY THE STATE IN
                            WHICH YOU TEACH?
GRADES/ASSESSMENTS:         N04, N08
CONDITIONING VAR LABEL:     TYP CERT
NAEP ID:                    T050001                     TOTAL NUMBER OF SPECIFIED CONTRASTS:     5
TYPE OF CONTRAST:           CLASS                       NUMBER OF INDEPENDENT CONTRASTS:         4

001 CERT-NO  (1          ) 0000TYPE OF TEACHING CERTIFICATION:  NONE
002 CERT-TMP (2          ) 1000TYPE OF TEACHING CERTIFICATION:  TEMPORARY, PROBATIONAL, PROVISIONAL, EMERGENCY
003 CERT-REG (3          ) 0100TYPE OF TEACHING CERTIFICATION:  REGULAR, BUT NOT HIGHEST
004 CERT-HGH (4          ) 0010TYPE OF TEACHING CERTIFICATION:  HIGHEST AVAILABLE
005 CERT-?   (M          ) 0001TYPE OF TEACHING CERTIFICATION:  MISSING, DOES NOT APPLY

CONDITIONING VARIABLE ID:  TCHR0016
DESCRIPTION:                WHAT IS THE HIGHEST ACADEMIC DEGREE YOU HOLD?
GRADES/ASSESSMENTS:         N04, N08
```

Table C-6 (continued)

Conditioning Variables Specific to the Geography Assessment

```
CONDITIONING VAR LABEL:    T_DEGREE
NAEP ID:                   T050101                   TOTAL NUMBER OF SPECIFIED CONTRASTS:     8
TYPE OF CONTRAST:          CLASS                     NUMBER OF INDEPENDENT CONTRASTS:         7

001 TDEG-HSD (1          ) 0000000                   TEACHER HIGHEST DEGREE:  HIGH SCHOOL DIPLOMA
002 TDEG-ASC (2          ) 1000000                   TEACHER HIGHEST DEGREE:  ASSOCIATES/VOCATIONAL
003 TDEG-BAC (3          ) 0100000                   TEACHER HIGHEST DEGREE:  BACHELOR'S HIGHEST DEGREE
004 TDEG-MAS (4          ) 0010000                   TEACHER HIGHEST DEGREE:  MASTER'S HIGHEST DEGREE
005 TDEG-EDS (5          ) 0001000                   TEACHER HIGHEST DEGREE:  EDUCATION SPECIALIST
006 TDEG-DOC (6          ) 0000100                   TEACHER HIGHEST DEGREE:  DOCTORATE
007 TDEG-PRO (7          ) 0000010                   TEACHER HIGHEST DEGREE:  PROFESSIONAL HIGHEST
                                                     DEGREE
008 TDEG-?   (M          ) 0000001                   TEACHER HIGHEST DEGREE:  MISSING, DOES NOT APPLY

CONDITIONING VARIABLE ID:  TCHR0017
DESCRIPTION:               TEACHER ENGLISH UNDERGRADUATE MAJOR
GRADES/ASSESSMENTS:        N04
CONDITIONING VAR LABEL:    UGRD ENG
NAEP ID:                   T040706                   TOTAL NUMBER OF SPECIFIED CONTRASTS:     2
TYPE OF CONTRAST:          CLASS                     NUMBER OF INDEPENDENT CONTRASTS:         1

001 UGRENG-Y (1          ) 0   TEACHER ENGLISH UNDERGRADUATE MAJOR:  YES
002 UGRENG-? (0,M        ) 1   TEACHER ENGLISH UNDERGRADUATE MAJOR:  MISSING, DOES NOT APPLY

CONDITIONING VARIABLE ID:  TCHR0018
DESCRIPTION:               TEACHER READING/LANGUAGE ARTS UNDERGRADUATE MAJOR
GRADES/ASSESSMENTS:        N04
CONDITIONING VAR LABEL:    UGRD RED
NAEP ID:                   T040707                   TOTAL NUMBER OF SPECIFIED CONTRASTS:     2
TYPE OF CONTRAST:          CLASS                     NUMBER OF INDEPENDENT CONTRASTS:         1

001 UGRRED-Y (1          ) 0   TEACHER READING UNDERGRADUATE MAJOR:  YES
002 UGRRED-? (0,M        ) 1   TEACHER READING UNDERGRADUATE MAJOR:  MISSING, DOES NOT APPLY

CONDITIONING VARIABLE ID:  TCHR0019
DESCRIPTION:               TEACHER ENGLISH GRADUATE MAJOR
GRADES/ASSESSMENTS:        N04
CONDITIONING VAR LABEL:    GRAD ENG
NAEP ID:                   T040807                   TOTAL NUMBER OF SPECIFIED CONTRASTS:     2
TYPE OF CONTRAST:          CLASS                     NUMBER OF INDEPENDENT CONTRASTS:         1

001 GRDENG-Y (1          ) 0   TEACHER ENGLISH GRADUATE MAJOR:       YES
002 GRDENG-? (0,M        ) 1   TEACHER ENGLISH GRADUATE MAJOR:       MISSING, DOES NOT APPLY

CONDITIONING VARIABLE ID:  TCHR0020
DESCRIPTION:               TEACHER READING/LANGUAGE ARTS GRADUATE MAJOR
GRADES/ASSESSMENTS:        N04
CONDITIONING VAR LABEL:    GRAD RED
NAEP ID:                   T040808                   TOTAL NUMBER OF SPECIFIED CONTRASTS:     2
TYPE OF CONTRAST:          CLASS                     NUMBER OF INDEPENDENT CONTRASTS:         1

001 GRDRED-Y (1          ) 0   TEACHER READING GRADUATE MAJOR:       YES
002 GRDRED-? (0,M        ) 1   TEACHER READING GRADUATE MAJOR:       MISSING, DOES NOT APPLY

CONDITIONING VARIABLE ID:  TCHR0021
DESCRIPTION:               HOW MUCH TIME HAVE YOU SPENT LAST YEAR IN READING DEVELOPMENT WORKSHOPS/SEMINARS?

GRADES/ASSESSMENTS:        N04
CONDITIONING VAR LABEL:    READ DEV
NAEP ID:                   T050201                   TOTAL NUMBER OF SPECIFIED CONTRASTS:     6
TYPE OF CONTRAST:          CLASS                     NUMBER OF INDEPENDENT CONTRASTS:         5
```

Table C-6 (continued)

Conditioning Variables Specific to the Geography Assessment

```
001 READDEV1 (1          ) 00000                  TIME I1 READING DEVELOPMENT WORKSHOPS:  NONE
002 READDEV2 (2          ) 10000                  TIME IN READING DEVELOPMENT WORKSHOPS:  LESS THAN
                                                  6 HOURS
003 READDEV3 (3          ) 01000                  TIME IN READING DEVELOPMENT WORKSHOPS:  6-15 HOURS
004 READDEV4 (4          ) 00100                  TIME IN READING DEVELOPMENT WORKSHOPS:  16-35
                                                  HOURS
005 READDEV5 (5          ) 00010                  TIME IN READING DEVELOPMENT WORKSHOPS:  MORE THAN
                                                  35 HOURS
006 READDEV5 (M          ) 00001                  TIME IN READING DEVELOPMENT WORKSHOPS:  MISSING,
                                                  DOES NOT APPLY

CONDITIONING VARIABLE ID:  TCHR0022
DESCRIPTION:               HOW MANY HOURS DO YOU HAVE DESIGNATED AS PREPERATION PERIODS PER WEEK?
GRADES/ASSESSMENTS:        N04, N08
CONDITIONING VAR LABEL:    PREP PER
NAEP ID:                   T051101                TOTAL NUMBER OF SPECIFIED CONTRASTS:      5
TYPE OF CONTRAST:          CLASS                  NUMBER OF INDEPENDENT CONTRASTS:          4

001 PRPER-0  (1          ) 0000TEACHER WEEKLY PREPARTION PERIODS:  NONE
002 PRPER-<1 (2          ) 1000TEACHER WEEKLY PREPARTION PERIODS:  LESS THAN 1
003 PRPER-12 (3          ) 0100TEACHER WEEKLY PREPARTION PERIODS:  1 TO 2
004 PRPER->2 (4          ) 0010TEACHER WEEKLY PREPARTION PERIODS:  MORE THAN 2
005 PRPER-?  (M          ) 0001TEACHER WEEKLY PREPARTION PERIODS:  MISSING, DOES NOT APPLY

CONDITIONING VARIABLE ID:  TCHR0023
DESCRIPTION:               ARE CURRICULUM SPECIALISTS AVAILABLE FOR READING?
GRADES/ASSESSMENTS:        N04
CONDITIONING VAR LABEL:    CURSPECS
NAEP ID:                   T041301                TOTAL NUMBER OF SPECIFIED CONTRASTS:      3
TYPE OF CONTRAST:          CLASS                  NUMBER OF INDEPENDENT CONTRASTS:          2

001 CSPECS-Y (1          ) 00  READING CURRICULUM SPECIALISTS:  YES
002 CSPECS-N (2          ) 10  READING CURRICULUM SPECIALISTS:  NO
003 CSPECS-? (M          ) 01  READING CURRICULUM SPECIALISTS:  MISSING, DOES NOT APPLY

CONDITIONING VARIABLE ID:  TCHR0024
DESCRIPTION:               TEACHER CERTIFICATION IN HISTORY
GRADES/ASSESSMENTS:        N04, N08
CONDITIONING VAR LABEL:    CERT HIS
NAEP ID:                   T040507                TOTAL NUMBER OF SPECIFIED CONTRASTS:      4
TYPE OF CONTRAST:          CLASS                  NUMBER OF INDEPENDENT CONTRASTS:          3

001 CERTH-Y  (1          ) 000 TEACHER HISTORY CERTIFICATION:  YES
002 CERTH-N  (2          ) 100 TEACHER HISTORY CERTIFICATION:  NO
003 CERTH-NS (3          ) 010 TEACHER HISTORY CERTIFICATION:  NOT OFFERED IN STATE
004 CERTH-?  (M          ) 001 TEACHER HISTORY CERTIFICATION:  MISSING, DOES NOT APPLY

CONDITIONING VARIABLE ID:  TCHR0025
DESCRIPTION:               TEACHER CERTIFICATION IN GEOGRAPHY
GRADES/ASSESSMENTS:        N04, N08
CONDITIONING VAR LABEL:    CERT GEO
NAEP ID:                   T040506                TOTAL NUMBER OF SPECIFIED CONTRASTS:      4
TYPE OF CONTRAST:          CLASS                  NUMBER OF INDEPENDENT CONTRASTS:          3

001 CERTG-Y  (1          ) 000 TEACHER GEOGRAPHY CERTIFICATION: YES
002 CERTG-N  (2          ) 100 TEACHER GEOGRAPHY CERTIFICATION: NO
003 CERTG-NS (3          ) 010 TEACHER GEOGRAPHY CERTIFICATION: NOT OFFERED IN STATE
004 CERTG-?  (M          ) 001 TEACHER GEOGRAPHY CERTIFICATION: MISSING, DOES NOT APPLY

CONDITIONING VARIABLE ID:  TCHR0026
DESCRIPTION:               TEACHER CERTIFICATION IN MIDDLE/JR. H.S. SOCIAL STUDIES
GRADES/ASSESSMENTS:        N08
```

Table C-6 (continued)

Conditioning Variables Specific to the Geography Assessment

```
CONDITIONING VAR LABEL:    CERT MSS
NAEP ID:                   T040509                          TOTAL NUMBER OF SPECIFIED CONTRASTS:     4
TYPE OF CONTRAST:          CLASS                            NUMBER OF INDEPENDENT CONTRASTS:         3

001 CERTM-Y  (1          ) 000 TEACHER MIDDLE/JR. H.S. SOCIAL STUDIES CERTIFICATION: YES
002 CERTM-N  (2          ) 100 TEACHER MIDDLE/JR. H.S. SOCIAL STUDIES CERTIFICATION: NO
003 CERTM-NS (3          ) 010 TEACHER MIDDLE/JR. H.S. SOCIAL STUDIES CERTIFICATION: NOT OFFERED IN STATE
004 CERTM-?  (M          ) 001 TEACHER MIDDLE/JR. H.S. SOCIAL STUDIES CERTIFICATION: MISSING, DOES NOT APPLY


CONDITIONING VARIABLE ID:  TCHR0027
DESCRIPTION:               TEACHER CERTIFICATION IN SECONDARY SOCIAL STUDIES TEACHER
GRADES/ASSESSMENTS:        N08
CONDITIONING VAR LABEL:    CERT SSS
NAEP ID:                   T040510                          TOTAL NUMBER OF SPECIFIED CONTRASTS:     4
TYPE OF CONTRAST:          CLASS                            NUMBER OF INDEPENDENT CONTRASTS:         3

001 CERTS-Y  (1          ) 000 TEACHER SECONDARY SOCIAL STUDIES TEACHER CERTIFICATION: YES
002 CERTS-N  (2          ) 100 TEACHER SECONDARY SOCIAL STUDIES TEACHER CERTIFICATION: NO
003 CERTS-NS (3          ) 010 TEACHER SECONDARY SOCIAL STUDIES TEACHER CERTIFICATION: NOT OFFERED IN STATE
004 CERTS-?  (M          ) 001 TEACHER SECONDARY SOCIAL STUDIES TEACHER CERTIFICATION: MISSING, DOES NOT APPLY

CONDITIONING VARIABLE ID:  TCHR0028
DESCRIPTION:               TEACHER CERTIFICATION IN OTHER
GRADES/ASSESSMENTS:        N04, N08
CONDITIONING VAR LABEL:    CERT OTH
NAEP ID:                   T040505                          TOTAL NUMBER OF SPECIFIED CONTRASTS:     4
TYPE OF CONTRAST:          CLASS                            NUMBER OF INDEPENDENT CONTRASTS:         3

001 CERTO-Y  (1          ) 000 TEACHER OTHER CERTIFICATION:    YES
002 CERTO-N  (2          ) 100 TEACHER OTHER CERTIFICATION:    NO
003 CERTO-NS (3          ) 010 TEACHER OTHER CERTIFICATION:    NOT OFFERED IN STATE
004 CERTO-?  (M          ) 001 TEACHER OTHER CERTIFICATION:    MISSING, DOES NOT APPLY

CONDITIONING VARIABLE ID:  TCHR0029
DESCRIPTION:               TEACHER UNDERGRADUATE MAJOR IN GEOGRAPHY
GRADES/ASSESSMENTS:        N04, N08
CONDITIONING VAR LABEL:    UGRD GEO
NAEP ID:                   T040708                          TOTAL NUMBER OF SPECIFIED CONTRASTS:     2
TYPE OF CONTRAST:          CLASS                            NUMBER OF INDEPENDENT CONTRASTS:         1

001 UGR GE-? (0,M        ) 0   TEACHER UNDERGRAD GEOGRAPHY MAJOR:  MISSING, DOES NOT APPLY
002 UGR GE-Y (1          ) 1   TEACHER UNDERGRAD GEOGRAPHY MAJOR:  YES

CONDITIONING VARIABLE ID:  TCHR0030
DESCRIPTION:               TEACHER UNDERGRADUATE MAJOR IN HISTORY
GRADES/ASSESSMENTS:        N04, N08
CONDITIONING VAR LABEL:    UGRD HIS
NAEP ID:                   T040709                          TOTAL NUMBER OF SPECIFIED CONTRASTS:     2
TYPE OF CONTRAST:          CLASS                            NUMBER OF INDEPENDENT CONTRASTS:         1

001 UGR HI-? (0,M        ) 0   TEACHER UNDERGRAD HISTORY MAJOR:    MISSING, DOES NOT APPLY
002 UGR HI-Y (1          ) 1   TEACHER UNDERGRAD HISTORY MAJOR:    YES

CONDITIONING VARIABLE ID:  TCHR0031
DESCRIPTION:               TEACHER UNDERGRADUATE MAJOR IN SOCIAL STUDIES EDUCATION
GRADES/ASSESSMENTS:        N04, N08
CONDITIONING VAR LABEL:    UGRD SS
NAEP ID:                   T040710                          TOTAL NUMBER OF SPECIFIED CONTRASTS:     2
TYPE OF CONTRAST:          CLASS                            NUMBER OF INDEPENDENT CONTRASTS:         1
```

BEST COPY AVAILABLE

Table C-6 (continued)

Conditioning Variables Specific to the Geography Assessment

```
001 UGR SS-? (0,M        ) 0   TEACHER UNDERGRAD SOCIAL STUDIES EDUCATION MAJOR: MISSING, DOES NOT APPLY
002 UGR SS-Y (1          ) 1   TEACHER UNDERGRAD SOCIAL STUDIES EDUCATION MAJOR: YES

CONDITIONING VARIABLE ID:  TCHR0032
DESCRIPTION:               TEACHER UNDERGRADUATE MAJOR IN OTHER
GRADES/ASSESSMENTS:        N04, N08
CONDITIONING VAR LABEL:    UGRD OTH
NAEP ID:                   T040705                       TOTAL NUMBER OF SPECIFIED CONTRASTS:     2
TYPE OF CONTRAST:          CLASS                         NUMBER OF INDEPENDENT CONTRASTS:         1

001 UGR OT-? (0,M        ) 0   TEACHER UNDERGRAD OTHER MAJOR:      MISSING, DOES NOT APPLY
002 UGR OT-Y (1          ) 1   TEACHER UNDERGRAD OTHER MAJOR:      YES

CONDITIONING VARIABLE ID:  TCHR0033
DESCRIPTION:               TEACHER GRADUATE MAJOR IN GEOGRAPHY
GRADES/ASSESSMENTS:        N04, N08
CONDITIONING VAR LABEL:    GRD GEO
NAEP ID:                   T040809                       TOTAL NUMBER OF SPECIFIED CONTRASTS:     2
TYPE OF CONTRAST:          CLASS                         NUMBER OF INDEPENDENT CONTRASTS:         1

001 GRD GE-? (0,M        ) 0   TEACHER GRADUATE GEOGRAPHY MAJOR:  MISSING, DOES NOT APPLY
002 GRD GE-Y (1          ) 1   TEACHER GRADUATE GEOGRAPHY MAJOR:  YES

CONDITIONING VARIABLE ID:  TCHR0034
DESCRIPTION:               TEACHER GRADUATE MAJOR IN HISTORY
GRADES/ASSESSMENTS:        N04, N08
CONDITIONING VAR LABEL:    GRD HIS
NAEP ID:                   T040810                       TOTAL NUMBER OF SPECIFIED CONTRASTS:     2
TYPE OF CONTRAST:          CLASS                         NUMBER OF INDEPENDENT CONTRASTS:         1

001 GRD HI-? (0,M        ) 0   TEACHER GRADUATE HISTORY MAJOR:    MISSING, DOES NOT APPLY
002 GRD HI-Y (1          ) 1   TEACHER GRADUATE HISTORY MAJOR:    YES

CONDITIONING VARIABLE ID:  TCHR0035
DESCRIPTION:               TEACHER GRADUATE MAJOR IN SOCIAL STUDIES EDUCATION
GRADES/ASSESSMENTS:        N04, N08
CONDITIONING VAR LABEL:    GRD SS
NAEP ID:                   T040811                       TOTAL NUMBER OF SPECIFIED CONTRASTS:     2
TYPE OF CONTRAST:          CLASS                         NUMBER OF INDEPENDENT CONTRASTS:         1

001 GRD SS-? (0,M        ) 0   TEACHER GRADUATE SOCIAL STUDIES EDUCATION MAJOR: MISSING, DOES NOT APPLY

002 GRD SS-Y (1          ) 1   TEACHER GRADUATE SOCIAL STUDIES EDUCATION MAJOR: YES

CONDITIONING VARIABLE ID:  TCHR0036
DESCRIPTION:               TEACHER GRADUATE MAJOR -OTHER
GRADES/ASSESSMENTS:        N04, N08
CONDITIONING VAR LABEL:    GRD OTH
NAEP ID:                   T040805                       TOTAL NUMBER OF SPECIFIED CONTRASTS:     2

TYPE OF CONTRAST:          CLASS                         NUMBER OF INDEPENDENT CONTRASTS:         1


001 GRA OT-? (0,M        ) 0   TEACHER GRADUATE OTHER MAJOR:  MISSING, DOES NOT APPLY
002 GRA OT-Y (1          ) 1   TEACHER GRADUATE OTHER MAJOR:  YES

CONDITIONING VARIABLE ID:  TCHR0037
DESCRIPTION:               DO YOU SAVE YOUR STUDENTS' SOCIAL STUDIES, HISTORY OR GEOGRAPHY WORK IN FOLDERS OR
                           PORTFOLIOS?
GRADES/ASSESSMENTS:        N04, N08
CONDITIONING VAR LABEL:    PORT-SS
NAEP ID:                   T050501                       TOTAL NUMBER OF SPECIFIED CONTRASTS:     3
```

Table C-6 (continued)

Conditioning Variables Specific to the Geography Assessment

```
TYPE OF CONTRAST:          CLASS                            NUMBER OF INDEPENDENT CONTRASTS:         2

001 PORT---Y (1          ) 00  PORTFOLIOS: YES
002 PORT---N (2          ) 10  PORTFOLIOS: NO
003 PORT---? (M          ) 01  PORTFOLIOS: MISSING

CONDITIONING VARIABLE ID:  TCHR0038
DESCRIPTION:               DO YOU SAVE YOUR STUDENTS' READING WORK IN FOLDERS OR PORTFOLIOS?
GRADES/ASSESSMENTS:        N04
CONDITIONING VAR LABEL:    PORT-RED
NAEP ID:                   T050601                          TOTAL NUMBER OF SPECIFIED CONTRASTS:     3
TYPE OF CONTRAST:          CLASS                            NUMBER OF INDEPENDENT CONTRASTS:         2

001 PORTR--Y (1          ) 00  READING PORTFOLIOS: YES
002 PORTR--N (2          ) 10  READING PORTFOLIOS: NO
003 PORTR--? (M          ) 01  READING PORTFOLIOS: MISSING

CONDITIONING VARIABLE ID:  TCHR0039
DESCRIPTION:               HOW MANY YEARS IN TOTAL HAVE YOU TAUGHT HISTORY.GEPGRAPHY, OR SOCIAL STUDIES? (4TH
                           GRADE)
GRADES/ASSESSMENTS:        N04
CONDITIONING VAR LABEL:    T4HISYRS
NAEP ID:                   T049902                          TOTAL NUMBER OF SPECIFIED CONTRASTS:     6
TYPE OF CONTRAST:          CLASS                            NUMBER OF INDEPENDENT CONTRASTS:         5

001 T4HISYR1 (1          ) 00000                            YEARS TEACHING HISTORY.GEPGRAPHY, OR SOCIAL
                                                            STUDIES: 2 YEARS OR LESS
002 T4HISYR2 (2          ) 10000                            YEARS TEACHING HISTORY.GEPGRAPHY, OR SOCIAL
                                                            STUDIES: 3-5 YEARS
003 T4HISYR3 (3          ) 01000                            YEARS TEACHING HISTORY.GEPGRAPHY, OR SOCIAL
                                                            STUDIES: 6-10 YEARS
004 T4HISYR4 (4          ) 00100                            YEARS TEACHING HISTORY.GEPGRAPHY, OR SOCIAL
                                                            STUDIES: 11-24 YEARS
005 T4HISYR5 (5          ) 00010                            YEARS TEACHING HISTORY.GEPGRAPHY, OR SOCIAL
                                                            STUDIES: 25 YEARS OR MORE
006 T4HISYR? (M          ) 00001                            YEARS TEACHING HISTORY.GEPGRAPHY, OR SOCIAL
                                                            STUDIES: MISSING, DOES NOT APPLY

CONDITIONING VARIABLE ID:  TCHR0040
DESCRIPTION:               DURING THE LAST YEAR IN GEOGRAPHY STAFF DEVELOPEMENT WORKSHOPS OR SEMINARS?
GRADES/ASSESSMENTS:        N04
CONDITIONING VAR LABEL:    TCHWRKSG
NAEP ID:                   T050401                          TOTAL NUMBER OF SPECIFIED CONTRASTS:     6
TYPE OF CONTRAST:          CLASS                            NUMBER OF INDEPENDENT CONTRASTS:         5

001 TWKSG-0  (1          ) 00000                            TEACHER WORKSHOPS OR SEMINARS IN GEOGRAPHY:  NONE
002 TWKSG-<6 (2          ) 10000                            TEACHER WORKSHOPS OR SEMINARS IN GEOGRAPHY:  LESS
                                                            THAN 6 HOURS
003 TWKSG-15 (3          ) 01000                            TEACHER WORKSHOPS OR SEMINARS IN GEOGRAPHY:  6 -
                                                            15 HOURS
004 TWKSG-35 (4          ) 00100                            TEACHER WORKSHOPS OR SEMINARS IN GEOGRAPHY:  16 -
                                                            35 HOURS
005 TWKSG>35 (5          ) 00010                            TEACHER WORKSHOPS OR SEMINARS IN GEOGRAPHY:  MORE
                                                            THAN 35 HOURS
006 TWKSG-?  (M          ) 00001                            TEACHER WORKSHOPS OR SEMINARS IN GEOGRAPHY:
                                                            MISSING, DOES NOT APPLY

CONDITIONING VARIABLE ID:  TCHR0041
DESCRIPTION:               ARE CURRICULUM SPECIALISTS AVAILABLE FOR GEOGRAPHY?
GRADES/ASSESSMENTS:        N04
CONDITIONING VAR LABEL:    CURSPECG
NAEP ID:                   T041303                          TOTAL NUMBER OF SPECIFIED CONTRASTS:     3
```

Table C-6 (continued)

Conditioning Variables Specific to the Geography Assessment

```
TYPE OF CONTRAST:          CLASS                           NUMBER OF INDEPENDENT CONTRASTS:          2

001 CSPECG-Y (1           ) 00  GEOGRAPHY CURRICULUM SPECIALISTS: YES
002 CSPECG-N (2           ) 10  GEOGRAPHY CURRICULUM SPECIALISTS: NO
003 CSPECG-? (M           ) 01  GEOGRAPHY CURRICULUM SPECIALISTS: MISSING, DOES NOT APPLY

CONDITIONING VARIABLE ID:  TCHR0042
DESCRIPTION:               ARE CURRICULUM SPECIALISTS AVAILABLE FOR SOCIAL STUDIES?
GRADES/ASSESSMENTS:        N04
CONDITIONING VAR LABEL:    CURSPECS
NAEP ID:                   T041304                         TOTAL NUMBER OF SPECIFIED CONTRASTS:      3
TYPE OF CONTRAST:          CLASS                           NUMBER OF INDEPENDENT CONTRASTS:          2

001 CSPECS-Y (1           ) 00  SOCIAL STUDIES CURRICULUM SPECIALISTS: YES
002 CSPECS-N (2           ) 10  SOCIAL STUDIES CURRICULUM SPECIALISTS: NO
003 CSPECS-? (M           ) 01  SOCIAL STUDIES CURRICULUM SPECIALISTS: MISSING, DOES NOT APPLY

CONDITIONING VARIABLE ID:  TCHR0043
DESCRIPTION:               ARE CURRICULUM SPECIALISTS AVAILABLE FOR HISTORY?
GRADES/ASSESSMENTS:        N04
CONDITIONING VAR LABEL:    CURSPECH
NAEP ID:                   T041305                         TOTAL NUMBER OF SPECIFIED CONTRASTS:      3
TYPE OF CONTRAST:          CLASS                           NUMBER OF INDEPENDENT CONTRASTS:          2

001 CSPECH-Y (1           ) 00  HISTORY CURRICULUM SPECIALISTS: YES
002 CSPECH-N (2           ) 10  HISTORY CURRICULUM SPECIALISTS: NO
003 CSPECH-? (M           ) 01  HISTORY CURRICULUM SPECIALISTS: MISSING, DOES NOT APPLY

CONDITIONING VARIABLE ID:  TSUB0001
DESCRIPTION:               IS CLASS ASSIGNED BY ABILITY?
GRADES/ASSESSMENTS:        N04, N08
CONDITIONING VAR LABEL:    CLASSABL
NAEP ID:                   T053201                         TOTAL NUMBER OF SPECIFIED CONTRASTS:      3
TYPE OF CONTRAST:          CLASS                           NUMBER OF INDEPENDENT CONTRASTS:          2

001 CLASAB-Y (1           ) 00  CLASS BY ABILITY: YES
002 CLASAB-N (2           ) 10  CLASS BY ABILITY: NO
003 CLASAB-? (M           ) 01  CLASS BY ABILITY: MISSING

CONDITIONING VARIABLE ID:  TSUB0002
DESCRIPTION:               WHAT IS THE ABLITITY LEVEL OF THE STUDENTS IN THIS CLASS?
GRADES/ASSESSMENTS:        N04, N08
CONDITIONING VAR LABEL:    CLABLLEV
NAEP ID:                   T053301                         TOTAL NUMBER OF SPECIFIED CONTRASTS:      5
TYPE OF CONTRAST:          CLASS                           NUMBER OF INDEPENDENT CONTRASTS:          4

001 CLABLV-H (1           ) 0000CLASS ABILITY LEVEL: PRIMARILY HIGH
002 CLASIZ-A (2           ) 1000CLASS ABILITY LEVEL: PRIMARILY AVERAGE
003 CLASIZ-L (3           ) 0100CLASS ABILITY LEVEL: PRIMARILY LOW
004 CLASIZ-M (4           ) 0010CLASS ABILITY LEVEL: PRIMARILY MIXED
005 CLASIZ-? (M           ) 0001CLASS ABILITY LEVEL: MISSING

CONDITIONING VARIABLE ID:  TSUB0003
DESCRIPTION:               ABOUT HOW MUCH TIME DO YOU SPEND WITH THIS CLASS ON GEOGRAPHY INSTRUCTION PER WEEK?
GRADES/ASSESSMENTS:        N04
CONDITIONING VAR LABEL:    AMT_TIME
NAEP ID:                   T053401                         TOTAL NUMBER OF SPECIFIED CONTRASTS:      5
TYPE OF CONTRAST:          CLASS                           NUMBER OF INDEPENDENT CONTRASTS:          4

001 TIME-30- (1           ) 0000GEOGRAPHY TIME: 30 MINUTES OR LESS
002 TIME-45M (2           ) 1000GEOGRAPHY TIME: 45 MINUTES
003 TIME-60M (3           ) 0100GEOGRAPHY TIME: 60 MINUTES
004 TIME-90+ (4           ) 0010GEOGRAPHY TIME: 90 MINUTES OR MORE
005 TIME-MIS (M           ) 0001GEOGRAPHY TIME: MISSING
```

Table C-6 (continued)

Conditioning Variables Specific to the Geography Assessment

```
CONDITIONING VARIABLE ID:  TSUB0004
DESCRIPTION:               HOW OFTEN DO YOU HAVE YOUR CLASS USE MAPS OR GLOBES?
GRADES/ASSESSMENTS:        N04, N08
CONDITIONING VAR LABEL:    USEMAPS
NAEP ID:                   T053501                          TOTAL NUMBER OF SPECIFIED CONTRASTS:     5
TYPE OF CONTRAST:          CLASS                            NUMBER OF INDEPENDENT CONTRASTS:         4

001 USEMAP-D (1          ) 0000HOW OFTEN USE MAPS OR GLOBES: ALMOST EVERY DAY
002 USEMAP-W (2          ) 1000HOW OFTEN USE MAPS OR GLOBES: ONCE OR TWICE A WEEK
003 USEMAP-M (3          ) 0100HOW OFTEN USE MAPS OR GLOBES: ONCE OR TWICE A MONTH
004 USEMAP-N (4          ) 0010HOW OFTEN USE MAPS OR GLOBES: NEVER OR HARDLY EVER
005 USEMAP-? (M          ) 0001HOW OFTEN USE MAPS OR GLOBES: MISSING

CONDITIONING VARIABLE ID:  TSUB0005
DESCRIPTION:               HOW OFTEN DO YOU HAVE YOUR CLASS MEASURE DISTANCES ON MAPS AND GLOBES?
GRADES/ASSESSMENTS:        N04, N08
CONDITIONING VAR LABEL:    DISTNT
NAEP ID:                   T053502                          TOTAL NUMBER OF SPECIFIED CONTRASTS:     5
TYPE OF CONTRAST:          CLASS                            NUMBER OF INDEPENDENT CONTRASTS:         4

001 DISTNT-D (1          ) 0000HOW OFTEN MEASURE DISTANCES: ALMOST EVERY DAY
002 DISTNT-W (2          ) 1000HOW OFTEN MEASURE DISTANCES: ONCE OR TWICE A WEEK
003 DISTNT-M (3          ) 0100HOW OFTEN MEASURE DISTANCES: ONCE OR TWICE A MONTH
004 DISTNT-N (4          ) 0010HOW OFTEN MEASURE DISTANCES: NEVER OR HARDLY EVER
005 DISTNT-? (M          ) 0001HOW OFTEN MEASURE DISTANCES: MISSING

CONDITIONING VARIABLE ID:  TSUB0006
DESCRIPTION:               HOW OFTEN DO YOU HAVE YOUR CLASS FIND PLACES USING LATITUDE AND LONGITUDE?
GRADES/ASSESSMENTS:        N04, N08
CONDITIONING VAR LABEL:    LATLONT
NAEP ID:                   T053503                          TOTAL NUMBER OF SPECIFIED CONTRASTS:     5
TYPE OF CONTRAST:          CLASS                            NUMBER OF INDEPENDENT CONTRASTS:         4

001 LATLNT-D (1          ) 0000HOW USE LATITUDE AND LONGITUDE: ALMOST EVERY DAY
002 LATLNT-W (2          ) 1000HOW USE LATITUDE AND LONGITUDE: ONCE OR TWICE A WEEK
003 LATLNT-M (3          ) 0100HOW USE LATITUDE AND LONGITUDE: ONCE OR TWICE A MONTH
004 LATLNT-N (4          ) 0010HOW USE LATITUDE AND LONGITUDE: NEVER OR HARDLY EVER
005 LATLNT-? (M          ) 0001HOW USE LATITUDE AND LONGITUDE: MISSING

CONDITIONING VARIABLE ID:  TSUB0007
DESCRIPTION:               HOW OFTEN DO YOU HAVE YOUR CLASS FIND CONTINENTS, OCEANS, RIVERS, AND MOUNTAINS?
GRADES/ASSESSMENTS:        N04, N08
CONDITIONING VAR LABEL:    CONTTCH
NAEP ID:                   T053504                          TOTAL NUMBER OF SPECIFIED CONTRASTS:     5
TYPE OF CONTRAST:          CLASS                            NUMBER OF INDEPENDENT CONTRASTS:         4

001 CONTCH-D (1          ) 0000FIND CONTINENTS: ALMOST EVERY DAY
002 CONTCH-W (2          ) 1000FIND CONTINETNS: ONCE OR TWICE A WEEK
003 CONTCH-M (3          ) 0100FIND CONTINENTS: ONCE OR TWICE A MONTH
004 CONTCH-N (4          ) 0010FIND CONTINENTS: NEVER OR HARDLY EVER
005 CONTCH-? (M          ) 0001FIND CONTINENTS: MISSING

CONDITIONING VARIABLE ID:  TSUB0008
DESCRIPTION:               HOW OFTEN DO YOU HAVE YOUR CLASS DISCUSS NATURAL RESOURCES?
GRADES/ASSESSMENTS:        N04, N08
CONDITIONING VAR LABEL:    NATREST
NAEP ID:                   T053505                          TOTAL NUMBER OF SPECIFIED CONTRASTS:     5
TYPE OF CONTRAST:          CLASS                            NUMBER OF INDEPENDENT CONTRASTS:         4

001 NATRST-1 (1          ) 0000NATURAL RESOURCES: ALMOST EVERY DAY
002 NATRST-2 (2          ) 1000NATURAL RESOURCES: ONCE OR TWICE A WEEK
003 NATRST-3 (3          ) 0100NATURAL RESOURCES: ONCE OR TWICE A DAY
004 NATRST-4 (4          ) 0010NATURAL RESOURCES: NEVER OR HARDLY EVER
005 NATRST-? (M          ) 0001NATURAL RESOURCES: MISSING
```

Table C-6 (continued)

Conditioning Variables Specific to the Geography Assessment

```
CONDITIONING VARIABLE ID:  TSUB0009
DESCRIPTION:               HOW OFTEN DO YOU HAVE YOUR CLASS DISCUSS WEATHER AND CLIMATE?
GRADES/ASSESSMENTS:        N04, N08
CONDITIONING VAR LABEL:    CLIMATT
NAEP ID:                   T053506                        TOTAL NUMBER OF SPECIFIED CONTRASTS:     5
TYPE OF CONTRAST:          CLASS                          NUMBER OF INDEPENDENT CONTRASTS:         4

001 CLIMTT-1 (1           ) 0000WEATHER AND CLIMATE: ALMOST EVERY DAY
002 CLIMTT-2 (2           ) 1000WEATHER AND CLIMATE: ONCE OR TWICE A WEEK
003 CLIMTT-3 (3           ) 0100WEATHER AND CLIMATE: ONCE OR TWICE A DAY
004 CLIMTT-4 (4           ) 0010WEATHER AND CLIMATE: NEVER OR HARDLY EVER
005 CLIMTT-? (M           ) 0001WEATHER AND CLIMATE: MISSING

CONDITIONING VARIABLE ID:  TSUB0010
DESCRIPTION:               HOW OFTEN DO YOU HAVE YOUR CLASS DISCUSS FOREIGN COUNTRIES AND CULTURES?
GRADES/ASSESSMENTS:        N04, N08
CONDITIONING VAR LABEL:    FORCOUN
NAEP ID:                   T053507                        TOTAL NUMBER OF SPECIFIED CONTRASTS:     5
TYPE OF CONTRAST:          CLASS                          NUMBER OF INDEPENDENT CONTRASTS:         4

001 FORCOU-1 (1           ) 0000FOREIGN COUNTRIES AND CULTURES: ALMOST EVERY DAY
002 FORCOU-2 (2           ) 1000FOREIGN COUNTRIES AND CULTURES: ONCE OR TWICE A WEEK
003 FORCOU-3 (3           ) 0100FOREIGN COUNTRIES AND CULTURES: ONCE OR TWICE A DAY
004 FORCOU-4 (4           ) 0010FOREIGN COUNTRIES AND CULTURES: NEVER OR HARDLY EVER
005 FORCOU-? (M           ) 0001FOREIGN COUNTRIES AND CULTURES: MISSING

CONDITIONING VARIABLE ID:  TSUB0011
DESCRIPTION:               HOW OFTEN DO YOU HAVE YOUR CLASS DISCUSS POPULATION GROWTH AND MOVEMENT?
GRADES/ASSESSMENTS:        N04, N08
CONDITIONING VAR LABEL:    POPGRO
NAEP ID:                   T053508                        TOTAL NUMBER OF SPECIFIED CONTRASTS:     5
TYPE OF CONTRAST:          CLASS                          NUMBER OF INDEPENDENT CONTRASTS:         4

001 POPGRO-1 (1           ) 0000POPULATION GROWTH AND MOVEMENT: ALMOST EVERY DAY
002 POPGRO-2 (2           ) 1000POPULATION GROWTH AND MOVEMENT: ONCE OR TWICE A WEEK
003 POPGRO-3 (3           ) 0100POPULATION GROWTH AND MOVEMENT: ONCE OR TWICE A DAY
004 POPGRO-4 (4           ) 0010POPULATION GROWTH AND MOVEMENT: NEVER OR HARDLY EVER
005 POPGRO-? (M           ) 0001POPULATION GROWTH AND MOVEMENT: MISSING

CONDITIONING VARIABLE ID:  TSUB0012
DESCRIPTION:               HOW OFTEN DO YOU HAVE YOUR CLASS DISCUSS CITIES, TOWNS, AND COMMUNITIES?
GRADES/ASSESSMENTS:        N04, N08
CONDITIONING VAR LABEL:    CITYTCH
NAEP ID:                   T053509                        TOTAL NUMBER OF SPECIFIED CONTRASTS:     5
TYPE OF CONTRAST:          CLASS                          NUMBER OF INDEPENDENT CONTRASTS:         4

001 CITYTC-1 (1           ) 0000CITIES, TOWNS, AND COMMUNITIES: ALMOST EVERY DAY
002 CITYTC-2 (2           ) 1000CITIES, TOWNS, AND COMMUNITIES: ONCE OR TWICE A WEEK
003 CITYTC-3 (3           ) 0100CITIES, TOWNS, AND COMMUNITIES: ONCE OR TWICE A DAY
004 CITYTC-4 (4           ) 0010CITIES, TOWNS, AND COMMUNITIES: NEVER OR HARDLY EVER
005 CITYTC-? (M           ) 0001CITIES, TOWNS, AND COMMUNITIES: MISSING

CONDITIONING VARIABLE ID:  TSUB0013
DESCRIPTION:               HOW OFTEN DO YOU HAVE YOUR CLASS DISCUSS ENVIRONMENTAL ISSUES?
GRADES/ASSESSMENTS:        N04, N08
CONDITIONING VAR LABEL:    ENVIRTC
NAEP ID:                   T053510                        TOTAL NUMBER OF SPECIFIED CONTRASTS:     5
TYPE OF CONTRAST:          CLASS                          NUMBER OF INDEPENDENT CONTRASTS:         4

001 ENVIRT-1 (1           ) 0000ENVIRONMENTAL ISSUES: ALMOST EVERY DAY
002 ENVIRT-2 (2           ) 1000ENVIRONMENTAL ISSUES: ONCE OR TWICE A WEEK
003 ENVIRT-3 (3           ) 0100ENVIRONMENTAL ISSUES: ONCE OR TWICE A DAY
004 ENVIRT-4 (4           ) 0010ENVIRONMENTAL ISSUES: NEVER OR HARDLY EVER
```

Table C-6 (continued)

Conditioning Variables Specific to the Geography Assessment

```
005 ENVIRT-? (M           ) 0001ENVIRONMENTAL ISSUES: MISSING

CONDITIONING VARIABLE ID:  TSUB0014
DESCRIPTION:               HOW OFTEN DO YOU HAVE YOUR CLASS DISCUSS REGIONS AND REGIONAL CONCEPTS?
GRADES/ASSESSMENTS:        N04, N08
CONDITIONING VAR LABEL:    REGTCH
NAEP ID:                   T053511                   TOTAL NUMBER OF SPECIFIED CONTRASTS:    5
TYPE OF CONTRAST:          CLASS                     NUMBER OF INDEPENDENT CONTRASTS:        4

001 REGTCH-1 (1           ) 0000REGIONS AND REGIONAL CONCEPTS: ALMOST EVERY DAY
002 REGTCH-2 (2           ) 1000REGIONS AND REGIONAL CONCEPTS: ONCE OR TWICE A WEEK
003 REGTCH-3 (3           ) 0100REGIONS AND REGIONAL CONCEPTS: ONCE OR TWICE A DAY
004 REGTCH-4 (4           ) 0010REGIONS AND REGIONAL CONCEPTS: NEVER OR HARDLY EVER
005 REGTCH-? (M           ) 0001REGIONS AND REGIONAL CONCEPTS: MISSING

CONDITIONING VARIABLE ID:  TSUB0015
DESCRIPTION:               ABOUT HOW MUCH GEOGRAPHY HOMEWORK DO YOU ASSIGN A STUDENT IN THIS CLASS EACH WEEK?
GRADES/ASSESSMENTS:        N04, N08
CONDITIONING VAR LABEL:    HISTHWK
NAEP ID:                   T053601                   TOTAL NUMBER OF SPECIFIED CONTRASTS:    6
TYPE OF CONTRAST:          CLASS                     NUMBER OF INDEPENDENT CONTRASTS:        5

001 HISHWK-N (1           ) 00000                    HOW MUCH GEOGRAPHY HOMEWORK: NONE
002 HISHWK-H (2           ) 10000                    HOW MUCH GEOGRAPHY HOMEWORK: 1/2 HOUR
003 HISHWK-1 (3           ) 01000                    HOW MUCH GEOGRAPHY HOMEWORK: 1 HOUR
004 HISHWK-2 (4           ) 00100                    HOW MUCH GEOGRAPHY HOMEWORK: 2 HOURS
005 HISHW-2+ (5           ) 00010                    HOW MUCH GEOGRAPHY HOMEWORK: MORE THAN 2 HOURS
006 HISHWK-? (M           ) 00001                    HOW MUCH GEOGRAPHY HOMEWORK: MISSING

CONDITIONING VARIABLE ID:  TSUB0016
DESCRIPTION:               HOW OFTEN DO YOU HAVE YOUR CLASS GO ON FIELD TRIPS, THAT RELATE TO THE STUDY OF
                           GEOGRAPHY?
GRADES/ASSESSMENTS:        N04, N08
CONDITIONING VAR LABEL:    TRIPS
NAEP ID:                   T053701                   TOTAL NUMBER OF SPECIFIED CONTRASTS:    4
TYPE OF CONTRAST:          CLASS                     NUMBER OF INDEPENDENT CONTRASTS:        3

001 TRIPS--O (1           ) 000 HOW OFTEN GO ON FIELD TRIPS: OFTEN
002 TRIPS--S (2           ) 100 HOW OFTEN GO ON FIELD TRIPS: SOMETIMES
003 TRIPS--N (3           ) 010 HOW OFTEN GO ON FIELD TRIPS: NEVER
004 TRIPS--? (M           ) 001 HOW OFTEN GO ON FIELD TRIPS: MISSING

CONDITIONING VARIABLE ID:  TSUB0017
DESCRIPTION:               HOW OFTEN DO YOU HAVE YOUR CLASS COMPLETE PROJECTS, THAT RELATE TO THE STUDY OF
                           GEOGRAPHY?
GRADES/ASSESSMENTS:        N04, N08
CONDITIONING VAR LABEL:    GEOPRJT
NAEP ID:                   T053801                   TOTAL NUMBER OF SPECIFIED CONTRASTS:    4
TYPE OF CONTRAST:          CLASS                     NUMBER OF INDEPENDENT CONTRASTS:        3

001 GEOPRJ-O (1           ) 000 HOW OFTEN COMPLETE PROJECTS: OFTEN
002 GEOPRJ-S (2           ) 100 HOW OFTEN COMPLETE PROJECTS: SOMETIMES
003 GEOPRJ-N (3           ) 010 HOW OFTEN COMPLETE PROJECTS: NEVER
004 GEOPRJ-? (M           ) 001 HOW OFTEN COMPLETE PROJECTS: MISSING

CONDITIONING VARIABLE ID:  TSUB0018
DESCRIPTION:               HOW OFTEN DO YOU HAVE YOUR CLASS PARTICIPATE IN LOCAL, STATE, OR NATIOAL GEOGRAPHY
                           CONTESTS?
GRADES/ASSESSMENTS:        N04, N08
CONDITIONING VAR LABEL:    GEOCNTS
NAEP ID:                   T053901                   TOTAL NUMBER OF SPECIFIED CONTRASTS:    4
TYPE OF CONTRAST:          CLASS                     NUMBER OF INDEPENDENT CONTRASTS:        3
```

Table C-6 (continued)

Conditioning Variables Specific to the Geography Assessment

```
001 GEOCNT-O (1          ) 000 HOW OFTEN PARTICIPATE IN GEOGRAPHY PROJECTS: OFTEN
002 GEOCNT-S (2          ) 100 HOW OFTEN PARTICIPATE IN GEOGRAPHY PROJECTS: SOMETIMES
003 GEOCNT-N (3          ) 010 HOW OFTEN PARTICIPATE IN GEOGRAPHY PROJECTS: NEVER
004 GEOCNT-? (M          ) 001 HOW OFTEN PARTICIPATE IN GEOGRAPHY PROJECTS: MISSING
CONDITIONING VARIABLE ID:  TSUB0019
DESCRIPTION:               HOW OFTEN DO YOU HAVE YOUR CLASS USE FILMS, VIDEOS, OR FILMSTRIPS IIN THE TEACHING
                           OF GEOGRAPHY?
GRADES/ASSESSMENTS:        N04, N08
CONDITIONING VAR LABEL:    VIDEOT
NAEP ID:                   T054001                    TOTAL NUMBER OF SPECIFIED CONTRASTS:     4
TYPE OF CONTRAST:          CLASS                      NUMBER OF INDEPENDENT CONTRASTS:         3

001 VIDEOT-O (1          ) 000 HOW OFTEN USE FILMS, VIDEOS, OR FILMSTRIPS: OFTEN
002 VIDEOT-S (2          ) 100 HOW OFTEN USE FILMS, VIDEOS, OR FILMSTRIPS: SOMETIMES
003 VIDEOT-N (3          ) 010 HOW OFTEN USE FILMS, VIDEOS, OR FILMSTRIPS: NEVER
004 VIDEOT-? (M          ) 001 HOW OFTEN USE FILMS, VIDEOS, OR FILMSTRIPS: MISSING

CONDITIONING VARIABLE ID:  TSUB0020
DESCRIPTION:               HOW OFTEN DO USE COMPUTERS IN TEACHING GEOGRAPHY?
GRADES/ASSESSMENTS:        N04, N08
CONDITIONING VAR LABEL:    USECOMT
NAEP ID:                   T054101                    TOTAL NUMBER OF SPECIFIED CONTRASTS:     4
TYPE OF CONTRAST:          CLASS                      NUMBER OF INDEPENDENT CONTRASTS:         3

001 USCOMT-O (1          ) 000 HOW OFTEN USE COMPUTERS: OFTEN
002 USCOMT-S (2          ) 100 HOW OFTEN USE COMPUTERS: SOMETIMES
003 USCOMT-N (3          ) 010 HOW OFTEN USE COMPUTERS: NEVER
004 USCOMT-? (M          ) 001 HOW OFTEN USE COMPUTERS: MISSING

CONDITIONING VARIABLE ID:  TSUB0021
DESCRIPTION:               HOW OFTEN DO YOU SEND OR TAKE YOUR CLASS TO THE LIBRARY?
GRADES/ASSESSMENTS:        N04, N08
CONDITIONING VAR LABEL:    GOLIBR
NAEP ID:                   T054201                    TOTAL NUMBER OF SPECIFIED CONTRASTS:     6
TYPE OF CONTRAST:          CLASS                      NUMBER OF INDEPENDENT CONTRASTS:         5

001 GOLIB--D (1          ) 00000                      HOW OFTEN TAKE TO LIBRARY: ALMOST EVERY DAY
002 GOLIB--W (2          ) 10000                      HOW OFTEN TAKE TO LIBRARY: ONCE OR TWICE A WEEK
003 GOLIB--M (3          ) 01000                      HOW OFTEN TAKE TO LIBRARY: ONCE OR TWICE A MONTH
004 GOLIB--N (4          ) 00100                      HOW OFTEN TAKE TO LIBRARY: NEVER OR HARDLY EVER
005 GOLIB-NO (5          ) 00010                      HOW OFTEN TAKE TO LIBRARY: THERE IS NO LIBRAY AT
                                                      THIS SCHOOL
006 GOLIBR-? (M          ) 00001                      HOW OFTEN TAKE TO LIBRARY: MISSING

CONDITIONING VARIABLE ID:  TSUB0022
DESCRIPTION:               HOW OFTEN DO YOU ASSIGN STUDENTS TO READ A BOOK FROM THE LIBRARY?
GRADES/ASSESSMENTS:        N04, N08
CONDITIONING VAR LABEL:    LIBBOOK
NAEP ID:                   T054202                    TOTAL NUMBER OF SPECIFIED CONTRASTS:     6
TYPE OF CONTRAST:          CLASS                      NUMBER OF INDEPENDENT CONTRASTS:         5

001 LIBOOK-D (1          ) 00000                      HOW OFTEN ASSIGN BOOK FROM LIBRARY: ALMOST EVERY
                                                      DAY
002 LIBOOK-W (2          ) 10000                      HOW OFTEN ASSIGN BOOK FROM LIBRARY: ONCE OR TWICE
                                                      A WEEK
003 LIBOOK-M (3          ) 01000                      HOW OFTEN ASSIGN BOOK FROM LIBRARY: ONCE OR TWICE
                                                      A MONTH
004 LIBOOK-N (4          ) 00100                      HOW OFTEN ASSIGN BOOK FROM LIBRARY: NEVER OR
                                                      HARDLY EVER
005 LIBOK-NO (5          ) 00010                      HOW OFTEN ASSIGN BOOK FROM LIBRARY: THERE IS NO
                                                      LIBRAY AT THIS SCHOOL
006 LIBOOK-? (M          ) 00001                      HOW OFTEN ASSIGN BOOK FROM LIBRARY: MISSING
```

Table C-6 (continued)

Conditioning Variables Specific to the Geography Assessment

```
CONDITIONING VARIABLE ID:  TSUB0023
DESCRIPTION:               WHICH BESTS DESCRIBES AVAILABILITY OF COMPUTERS FOR USE BY STUDENTS IN YOUR
                           GEOGRAPHY CLASS?
GRADES/ASSESSMENTS:        N04, N08
CONDITIONING VAR LABEL:    COMPAVA
NAEP ID:                   T054301                    TOTAL NUMBER OF SPECIFIED CONTRASTS:     4
TYPE OF CONTRAST:          CLASS                      NUMBER OF INDEPENDENT CONTRASTS:         3

001 COMPAV-N (1          ) 000 CPMPUTERS NOT AVAILABLE
002 COMPAV-D (2          ) 100 COMPUTERS AVAILABLE, BUT DIFFICULT TO USE.
003 COMPAV-C (3          ) 010 COMPUTERS AVAILABLE IN CLASSROOM.
004 COMPAV-? (M          ) 001 AVAILABILITY OF COMPUTERS ARE MISSING

CONDITIONING VARIABLE ID:  TSUB0024
DESCRIPTION:               HOW OFTEN DO YOU USE MULTIPLE CHOICE TESTS TO ASSESS STUDENT PROGRESS IN GEOGRAPHY?
GRADES/ASSESSMENTS:        N04, N08
CONDITIONING VAR LABEL:    MCTESTS
NAEP ID:                   T054401                    TOTAL NUMBER OF SPECIFIED CONTRASTS:     5
TYPE OF CONTRAST:          CLASS                      NUMBER OF INDEPENDENT CONTRASTS:         4

001 MCTEST-W (1          ) 0000USE MULTIPLE CHOICE TESTS:  ONCE OR TWICE A WEEK
002 MCTEST-M (2          ) 1000USE MULTIPLE CHOICE TESTS:  ONCE OR TWICE A MONTH
003 MCTEST-Y (3          ) 0100USE MULTIPLE CHOICE TESTS:  ONCE OR TWICE A YEAR
004 MCTEST-N (4          ) 0010USE MULTIPLE CHOICE TESTS:  NEVER OF HARDLY EVER
005 MCTEST-? (M          ) 0001USE MULTIPLE CHOICE TESTS:  MISSING

CONDITIONING VARIABLE ID:  TSUB0025
DESCRIPTION:               HOW OFTEN DO YOU USE SHORT OR LONG WRITTEN RESPONSE TO ASSESS STUDENT PROGRESS IN
                           GEOGRAPHY?
GRADES/ASSESSMENTS:        N04, N08
CONDITIONING VAR LABEL:    WRTRESP
NAEP ID:                   T054402                    TOTAL NUMBER OF SPECIFIED CONTRASTS:     5
TYPE OF CONTRAST:          CLASS                      NUMBER OF INDEPENDENT CONTRASTS:         4

001 WRTRSP-W (1          ) 0000USE SHORT OR LONG WRITTEN RESPONSES: ONCE OR TWICE A WEEK
002 WRTRSP-M (2          ) 1000USE SHORT OR LONG WRITTEN RESPONSES: ONCE OR TWICE A MONTH
003 WRTRSP-Y (3          ) 0100USE SHORT OR LONG WRITTEN RESPONSES: ONCE OR TWICE A YEAR
004 WRTRSP-N (4          ) 0010USE SHORT OR LONG WRITTEN RESPONSES: NEVER OF HARDLY EVER
005 WRTRSP-? (M          ) 0001USE SHORT OR LONG WRITTEN RESPONSES: MISSING

CONDITIONING VARIABLE ID:  TSUB0026
DESCRIPTION:               HOW OFTEN DO YOU USE INDIVIDUAL OR GROUP PROJECTS TO ASSESS STUDENT PROGRESS IN
                           GEOGRAPHY?
GRADES/ASSESSMENTS:        N04, N08
CONDITIONING VAR LABEL:    GRPPROJ
NAEP ID:                   T054403                    TOTAL NUMBER OF SPECIFIED CONTRASTS:     5
TYPE OF CONTRAST:          CLASS                      NUMBER OF INDEPENDENT CONTRASTS:         4

001 GRPPRJ-W (1          ) 0000USE PROJECTS: ONCE OR TWICE A WEEK
002 GRPPRJ-M (2          ) 1000USE PROJECTS: ONCE OR TWICE A MONTH
003 GRPPRJ-Y (3          ) 0100USE PROJECTS: ONCE OR TWICE A YEAR
004 GRPPRJ-N (4          ) 0010USE PROJECTS: NEVER OF HARDLY EVER
005 GRPPRJ-? (M          ) 0001USE PROJECTS: MISSING

CONDITIONING VARIABLE ID:  TSUB0027
DESCRIPTION:               HOW OFTEN DO YOU USE PORTFOLIO COLLECTIONS TO ASSESS STUDENT PROGRESS IN GEOGRAPHY?
GRADES/ASSESSMENTS:        N04, N08
CONDITIONING VAR LABEL:    TPORTF
NAEP ID:                   T054404                    TOTAL NUMBER OF SPECIFIED CONTRASTS:     5
TYPE OF CONTRAST:          CLASS                      NUMBER OF INDEPENDENT CONTRASTS:         4

001 TPORT--W (1          ) 0000USE PORTFOLIOS: ONCE OR TWICE A WEEK
002 TPORT--M (2          ) 1000USE PORTFOLIOS: ONCE OR TWICE A MONTH
003 TPORT--Y (3          ) 0100USE PORTFOLIOS: ONCE OR TWICE A YEAR
004 TPORT--N (4          ) 0010USE PORTFOLIOS: NEVER OF HARDLY EVER
```

BEST COPY AVAILABLE

Table C-6 (continued)

Conditioning Variables Specific to the Geography Assessment

```
005 TPORT--? (M          ) 0001USE PORTFOLIOS: MISSING
CONDITIONING VARIABLE ID:  TSUB0028
DESCRIPTION:               WHAT IS YOUR AVERAGE GEOGRAPHY CLASS SIZE?
GRADES/ASSESSMENTS:        N04
CONDITIONING VAR LABEL:    CLASSSZ4
NAEP ID:                   T054501                TOTAL NUMBER OF SPECIFIED CONTRASTS:     6
TYPE OF CONTRAST:          CLASS                  NUMBER OF INDEPENDENT CONTRASTS:         5

001 CL401-20 (1          ) 00000                  CLASS SIZE:  1-20 STUDENTS    (FOURTH GRADE)
002 CL421-25 (2          ) 10000                  CLASS SIZE: 21-25 STUDENTS    (FOURTH GRADE)
003 CL426-30 (3          ) 01000                  CLASS SIZE: 26-30 STUDENTS    (FOURTH GRADE)
004 CL431-35 (4          ) 00100                  CLASS SIZE: 31-35 STUDENTS    (FOURTH GRADE)
005 CL436+   (5          ) 00010                  CLASS SIZE: 36 OR MORE STUDENTS  (FOURTH GRADE)
006 CL4-MISS (M          ) 00001                  CLASS SIZE: MISSING           (FOURTH GRADE)

CONDITIONING VARIABLE ID:  TSUB0026
DESCRIPTION:               WHAT IS THE PRIMARY EMPHASIS OF THIS COURSE?
GRADES/ASSESSMENTS:        N08
CONDITIONING VAR LABEL:    PEMPH
NAEP ID:                   T055501                TOTAL NUMBER OF SPECIFIED CONTRASTS:     5
TYPE OF CONTRAST:          CLASS                  NUMBER OF INDEPENDENT CONTRASTS:         4

001 PEMP-HIS (1          ) 0000PRIMARY EMPHASIS: GEOGRAPHY
002 PEMP-GEO (2          ) 1000PRIMARY EMPHASIS: HISTORY
003 PEMP-SSO (3          ) 0100PRIMARY EMPHASIS: SOCIAL STUDIES
004 PEMP-OTH (4          ) 0010PRIMARY EMPHASIS: OTHER
005 PEMP-?   (M          ) 0001PRIMARY EMPHASIS: MISSING

CONDITIONING VARIABLE ID:  TSUB0030
DESCRIPTION:               WHAT IS YOUR GEOGRAPHY CLASS SIZE?
GRADES/ASSESSMENTS:        N08
CONDITIONING VAR LABEL:    CLASSSZ8
NAEP ID:                   T055701                TOTAL NUMBER OF SPECIFIED CONTRASTS:     6
TYPE OF CONTRAST:          CLASS                  NUMBER OF INDEPENDENT CONTRASTS:         5

001 CL801-20 (1          ) 00000                  CLASS SIZE:  1-20 STUDENTS     (EIGHTH GRADE)
002 CL821-25 (2          ) 10000                  CLASS SIZE: 21-25 STUDENTS     (EIGHTH GRADE)
003 CL826-30 (3          ) 01000                  CLASS SIZE: 26-30 STUDENTS     (EIGHTH GRADE)
004 CL831-35 (4          ) 00100                  CLASS SIZE: 31-35 STUDENTS     (EIGHTH GRADE)
005 CL836+   (5          ) 00010                  CLASS SIZE: 36 OR MORE STUDENTS (EIGHTH GRADE)
006 CL8-MISS (M          ) 00001                  CLASS SIZE: MISSING             (EIGHTH GRADE)

CONDITIONING VARIABLE ID:  TSUB0031
DESCRIPTION:               HOW WELL PREPARED DO YOU FEEL YOU ARE TO TEACH GEOGRAPHY AT THE ELEMENTARY-SCHOOL
                           LEVEL?
GRADES/ASSESSMENTS:        N04
CONDITIONING VAR LABEL:    PREPGEOG
NAEP ID:                   T052801                TOTAL NUMBER OF SPECIFIED CONTRASTS:     5
TYPE OF CONTRAST:          CLASS                  NUMBER OF INDEPENDENT CONTRASTS:         4

001 PRPGEO-V (1          ) 0000PREPARED TO TEACH GEOGRAPHY: VERY PREPARED
002 PRPGEO-A (2          ) 1000PREPARED TO TEACH GEOGRAPHY: ADEQUATELY PREPARED
003 PRPGEO-S (3          ) 0100PREPARED TO TEACH GEOGRAPHY: SOMEWHAT PREPARED
004 PRPGEO-U (4          ) 0010PREPARED TO TEACH GEOGRAPHY: UNPREPARED
005 PRPGEO-? (M          ) 0001PREPARED TO TEACH GEOGRAPHY: MISSING

CONDITIONING VARIABLE ID:  TSUB0032
DESCRIPTION:               HOW WELL PREPARED DO YOU FEEL YOU ARE TO TEACH HISTORY AT THE ELEMENTARY-SCHOOL
                           LEVEL?
GRADES/ASSESSMENTS:        N04
CONDITIONING VAR LABEL:    PREPHIST
NAEP ID:                   T052802                TOTAL NUMBER OF SPECIFIED CONTRASTS:     5
TYPE OF CONTRAST:          CLASS                  NUMBER OF INDEPENDENT CONTRASTS:         4
```

Table C-6 (continued)

Conditioning Variables Specific to the Geography Assessment

```
001 PRPHIS-V (1          ) 0000PREPARED TO TEACH HISTORY: VERY PREPARED
002 PRPHIS-A (2          ) 1000PREPARED TO TEACH HISTORY: ADEQUATELY PREPARED
003 PRPHIS-S (3          ) 0100PREPARED TO TEACH HISTORY: SOMEWHAT PREPARED
004 PRPHIS-U (4          ) 0010PREPARED TO TEACH HISTORY: UNPREPARED
005 PRPHIS-? (M          ) 0001PREPARED TO TEACH HISTORY: MISSING

CONDITIONING VARIABLE ID:  TSUB0033
DESCRIPTION:               HOW WELL PREPARED DO YOU FEEL YOU ARE TO TEACH SOCIAL STUDIES AT THE
                           ELEMENTARY-SCHOOL LEVEL?
GRADES/ASSESSMENTS:        N04
CONDITIONING VAR LABEL:    PREPSOCS
NAEP ID:                   T052803                  TOTAL NUMBER OF SPECIFIED CONTRASTS:     5
TYPE OF CONTRAST:          CLASS                    NUMBER OF INDEPENDENT CONTRASTS:         4

001 PRPSS--V (1          ) 0000PREPARED TO TEACH SOCIAL STUDIES: VERY PREPARED
002 PRPSS--A (2          ) 1000PREPARED TO TEACH SOCIAL STUDIES: ADEQUATELY PREPARED
003 PRPSS--S (3          ) 0100PREPARED TO TEACH SOCIAL STUDIES: SOMEWHAT PREPARED
004 PRPSS--U (4          ) 0010PREPARED TO TEACH SOCIAL STUDIES: UNPREPARED
005 PRPSS--? (M          ) 0001PREPARED TO TEACH SOCIAL STUDIES: MISSING

CONDITIONING VARIABLE ID:  TSUB0034
DESCRIPTION:               DURING THE LAST FIVE YEARS, HAVE YOU TAKEN ANY OF THE FOLLOWING RELATED TO
                           GEOGRAPHY?
GRADES/ASSESSMENTS:        N04, N08
CONDITIONING VAR LABEL:    GEOGEDU
NAEP ID:                   T052901                  TOTAL NUMBER OF SPECIFIED CONTRASTS:     5
TYPE OF CONTRAST:          CLASS                    NUMBER OF INDEPENDENT CONTRASTS:         4

001 GEOEDU-V (1          ) 0000TAKEN RELATED TO GEOGRAPHY: UNIVERSITY COURSES
002 GEOEDU-A (2          ) 1000TAKEN RELATED TO GEOGRAPHY: PROFESSIONAL DEVELOPMENT WORKSHOPS OR SEMINARS
003 GEOEDU-S (3          ) 0100TAKEN RELATED TO GEOGRAPHY: OTHER PROFESSIONAL ACTIVIES
004 GEOEDU-U (4          ) 0010TAKEN RELATED TO GEOGRAPHY: NONE
005 GEOEDU-? (M          ) 0001TAKEN RELATED TO GEOGRAPHY: MISSING

CONDITIONING VARIABLE ID:  TSUB0035
DESCRIPTION:               DURING THE LAST FIVE YEARS, HAVE YOU TAKEN ANY OF THE FOLLOWING RELATED TO HISTORY?
GRADES/ASSESSMENTS:        N04, N08
CONDITIONING VAR LABEL:    HISTEDU
NAEP ID:                   T052902                  TOTAL NUMBER OF SPECIFIED CONTRASTS:     5
TYPE OF CONTRAST:          CLASS                    NUMBER OF INDEPENDENT CONTRASTS:         4

001 HISEDU-V (1          ) 0000TAKEN RELATED TO HISTORY: UNIVERSITY COURSES
002 HISEDU-A (2          ) 1000TAKEN RELATED TO HISTORY: PROFESSIONAL DEVELOPMENT WORKSHOPS OR SEMINARS
003 HISEDU-S (3          ) 0100TAKEN RELATED TO HISTORY: OTHER PROFESSIONAL ACTIVIES
004 HISEDU-U (4          ) 0010TAKEN RELATED TO HISTORY: NONE
005 HISEDU-? (M          ) 0001TAKEN RELATED TO HISTORY: MISSING

CONDITIONING VARIABLE ID:  TSUB0036
DESCRIPTION:               DURING THE LAST FIVE YEARS, HAVE YOU TAKEN ANY OF THE FOLLOWING RELATED TO OTHER
                           SOCIAL STUDIES?
GRADES/ASSESSMENTS:        N04, N08
CONDITIONING VAR LABEL:    OSSEDU
NAEP ID:                   T052903                  TOTAL NUMBER OF SPECIFIED CONTRASTS:     5
TYPE OF CONTRAST:          CLASS                    NUMBER OF INDEPENDENT CONTRASTS:         4

001 OSSEDU-V (1          ) 0000TAKEN RELATED TO OTHER SOCIAL STUDIES: UNIVERSITY COURSES
002 OSSEDU-A (2          ) 1000TAKEN RELATED TO OTHER SOCIAL STUDIES: PROFESSIONAL DEVELOPMENT WORKSHOPS OR
                           SEMINARS
003 OSSEDU-S (3          ) 0100TAKEN RELATED TO OTHER SOCIAL STUDIES: OTHER PROFESSIONAL ACTIVIES
004 OSSEDU-U (4          ) 0010TAKEN RELATED TO OTHER SOCIAL STUDIES: NONE
005 OSSEDU-? (M          ) 0001TAKEN RELATED TO OTHER SOCIAL STUDIES: MISSING

CONDITIONING VARIABLE ID:  TSUB0037
DESCRIPTION:               DURING THE LAST FIVE YEARS, HAVE YOU TAKEN ANY OF THE FOLLOWING RELATED TO SCIENCE?
```

BEST COPY AVAILABLE

Table C-6 (continued)

Conditioning Variables Specific to the Geography Assessment

```
GRADES/ASSESSMENTS:          N04
CONDITIONING VAR LABEL:      SCIEDU
NAEP ID:                     T052904                      TOTAL NUMBER OF SPECIFIED CONTRASTS:     5
TYPE OF CONTRAST:            CLASS                        NUMBER OF INDEPENDENT CONTRASTS:         4


001 SCIEDU-V (1            ) 0000TAKEN RELATED TO SCIENCE: UNIVERSITY COURSES
002 SCIEDU-A (2            ) 1000TAKEN RELATED TO SCIENCE: PROFESSIONAL DEVELOPMENT WORKSHOPS OR SEMINARS
003 SCIEDU-S (3            ) 0100TAKEN RELATED TO SCIENCE: OTHER PROFESSIONAL ACTIVIES
004 SCIEDU-U (4            ) 0010TAKEN RELATED TO SCIENCE: NONE
005 SCIEDU-? (M            ) 0001TAKEN RELATED TO SCIENCE: MISSING

CONDITIONING VARIABLE ID:    TSUB0038
DESCRIPTION:                 WHICH OF THE FOLLOWING PLAYS THE CENTRAL ROLE IN YOUR GEOGRAPHY CURRICULUM?
GRADES/ASSESSMENTS:          N04
CONDITIONING VAR LABEL:      CRLECUR
NAEP ID:                     T053001                      TOTAL NUMBER OF SPECIFIED CONTRASTS:     7
TYPE OF CONTRAST:            CLASS                        NUMBER OF INDEPENDENT CONTRASTS:         6

001 CRLE-TXT (1            ) 000000                       CENTRAL ROLE IN CURRICULUM:  TEXTBOOK-ESTABLISHED
                                                          CURRICULUM GUIDES
002 CRLE-DST (2            ) 100000                       CENTRAL ROLE IN CURRICULUM:  DISTRICT-DEVELOPED
                                                          CURRICULUM GUIDES
003 CRLE-SCH (3            ) 010000                       CENTRAL ROLE IN CURRICULUM:  SCHOOL-DEVELOPED
                                                          CURRICULUM GUIDES
004 CRLE-IND (4            ) 001000                       CENTRAL ROLE IN CURRICULUM:  INDIVIDUALLY
                                                          DEVELOPED CURRICULUM GUIDES
005 CRLE-STA (5            ) 000100                       CENTRAL ROLE IN CURRICULUM:  STATE CURRICULUM
                                                          GUIDES OR FRAMEWORKS
006 CRLE-EXA (6            ) 000010                       CENTRAL ROLE IN CURRICULUM:  PUBLISHED EXAMS OR
                                                          ASSESSMENTS
007 CRLE-?   (M            ) 000001                       CENTRAL ROLE IN CURRICULUM:  MISSING

CONDITIONING VARIABLE ID:    TSUB0039
DESCRIPTION:                 WHICH OF THE FOLLOWING PLAYS THE CENTRAL ROLE IN YOUR GEOGRAPHY CURRICULUM?
GRADES/ASSESSMENTS:          N08
CONDITIONING VAR LABEL:      CRLECU8
NAEP ID:                     T053002                      TOTAL NUMBER OF SPECIFIED CONTRASTS:     6
TYPE OF CONTRAST:            CLASS                        NUMBER OF INDEPENDENT CONTRASTS:         5

001 CRL8-TXT (1            ) 00000                        CENTRAL ROLE IN CURRICULUM:  TEXTBOOK-ESTABLISHED
                                                          CURRICULUM GUIDES   (EIGHTH GRADE)
002 CRL8-DST (2            ) 10000                        CENTRAL ROLE IN CURRICULUM:  DISTRICT-DEVELOPED
                                                          CURRICULUM GUIDES      (EIGHTH GRADE)
003 CRL8-SCH (3            ) 01000                        CENTRAL ROLE IN CURRICULUM:  SCHOOL-DEVELOPED
                                                          CURRICULUM GUIDES        (EIGHTH GRADE)
004 CRL8-IND (4            ) 00100                        CENTRAL ROLE IN CURRICULUM:  INDIVIDUALLY
                                                          DEVELOPED CURRICULUM GUIDES (EIGHTH GRADE)
005 CRL8-EXA (5            ) 00010                        CENTRAL ROLE IN CURRICULUM:  PUBLISHED EXAMS OR
                                                          ASSESSMENTS            (EIGHTH GRADE)
006 CRL8-?   (M            ) 00001                        CENTRAL ROLE IN CURRICULUM:  MISSING

CONDITIONING VARIABLE ID:    TSUB0040
DESCRIPTION:                 HOW MUCH DO YOU LIKE TEACHING GEOGRAPHY?
GRADES/ASSESSMENTS:          N04
CONDITIONING VAR LABEL:      LKTCHGE
NAEP ID:                     T053101                      TOTAL NUMBER OF SPECIFIED CONTRASTS:     4
TYPE OF CONTRAST:            CLASS                        NUMBER OF INDEPENDENT CONTRASTS:         3

001 LKGEO-FV (1            ) 000 GEOGRAPHY: IS ONE OF MY FAVORITE SUBJECTS TO TEACH.
002 LKGEO-OT (2            ) 100 GEOGRAPHY: I LIKE TEACHING OTHER SUBJECTS BETTER.
003 LKGEO-NO (3            ) 010 GEOGRAPHY: I DO NOT ENJOY TEACHING UNITED STATES HISTORY.
004 LKGEO-?  (M            ) 001 GEOGRAPHY: MISSING
```

Table C-6 (continued)

Conditioning Variables Specific to the Geography Assessment

```
CONDITIONING VARIABLE ID:  TSUB0041
DESCRIPTION:               HOW WELL PREPARED DO YOU FEEL YOU ARE TO TEACH GEOGRAPHY AT THE JUNIOR HIGH LEVEL?
GRADES/ASSESSMENTS:        N08
CONDITIONING VAR LABEL:    PREPGEO
NAEP ID:                   T055301                         TOTAL NUMBER OF SPECIFIED CONTRASTS:     5
TYPE OF CONTRAST:          CLASS                           NUMBER OF INDEPENDENT CONTRASTS:         4

001 PRPGEO-V (1          ) 0000PREPARED TO TEACH JR. HI GEOGRAPHY: VERY PREPARED
002 PRPGEO-A (2          ) 1000PREPARED TO TEACH JR. HI GEOGRAPHY: ADEQUATELY PREPARED
003 PRPGEO-S (3          ) 0100PREPARED TO TEACH JR. HI GEOGRAPHY: SOMEWHAT PREPARED
004 PRPGEO-U (4          ) 0010PREPARED TO TEACH JR. HI GEOGRAPHY: UNPREPARED
005 PRPGEO-? (M          ) 0001PREPARED TO TEACH JR. HI GEOGRAPHY: MISSING

CONDITIONING VARIABLE ID:  TSUB0042
DESCRIPTION:               HOW WELL PREPARED DO YOU FEEL YOU ARE TO TEACH HISTORY AT THE JUNIOR HIGH LEVEL?
GRADES/ASSESSMENTS:        N08
CONDITIONING VAR LABEL:    PREPHIS
NAEP ID:                   T055302                         TOTAL NUMBER OF SPECIFIED CONTRASTS:     5
TYPE OF CONTRAST:          CLASS                           NUMBER OF INDEPENDENT CONTRASTS:         4

001 PRPHIS-V (1          ) 0000PREPARED TO TEACH JR. HI HISTORY: VERY PREPARED
002 PRPHIS-A (2          ) 1000PREPARED TO TEACH JR. HI HISTORY: ADEQUATELY PREPARED
003 PRPHIS-S (3          ) 0100PREPARED TO TEACH JR. HI HISTORY: SOMEWHAT PREPARED
004 PRPHIS-U (4          ) 0010PREPARED TO TEACH JR. HI HISTORY: UNPREPARED
005 PRPHIS-? (M          ) 0001PREPARED TO TEACH JR. HI HISTORY: MISSING

CONDITIONING VARIABLE ID:  TSUB0043
DESCRIPTION:               HOW WELL PREPARED DO YOU FEEL YOU ARE TO TEACH OTHER SOCIAL STUDIES AT THE JUNIOR
                           HIGH LEVEL?
GRADES/ASSESSMENTS:        N08
CONDITIONING VAR LABEL:    PREPOSS
NAEP ID:                   T055303                         TOTAL NUMBER OF SPECIFIED CONTRASTS:     5
TYPE OF CONTRAST:          CLASS                           NUMBER OF INDEPENDENT CONTRASTS:         4

001 PRPOSS-V (1          ) 0000PREPARED TO TEACH JR. HI OTHER SOCIAL STUDIES: VERY PREPARED
002 PRPOSS-A (2          ) 1000PREPARED TO TEACH JR. HI OTHER SOCIAL STUDIES: ADEQUATELY PREPARED
003 PRPOSS-S (3          ) 0100PREPARED TO TEACH JR. HI OTHER SOCIAL STUDIES: SOMEWHAT PREPARED
004 PRPOSS-U (4          ) 0010PREPARED TO TEACH JR. HI OTHER SOCIAL STUDIES: UNPREPARED
005 PRPOSS-? (M          ) 0001PREPARED TO TEACH JR. HI OTHER SOCIAL STUDIES: MISSING

CONDITIONING VARIABLE ID:  TSUB0044
DESCRIPTION:               HOW MUCH DO YOU LIKE TEACHING GEOGRAPHY?
GRADES/ASSESSMENTS:        N08
CONDITIONING VAR LABEL:    LKTCHG8
NAEP ID:                   T055401                         TOTAL NUMBER OF SPECIFIED CONTRASTS:     5
TYPE OF CONTRAST:          CLASS                           NUMBER OF INDEPENDENT CONTRASTS:         4

001 LKGE8-FV (1          ) 0000GEOGRAPHY: IS ONE OF MY FAVORITE SUBJECTS TO TEACH.   (EIGHTH GRADE)
002 LKGE8-IN (2          ) 1000GEOGRAPHY: I FIND IT INTERESTING TO TEACH.            (EIGHTH GRADE)
003 LKGE8-OT (3          ) 0100GEOGRAPHY: I LIKE TEACHING OTHER SUBJECTS BETTER.     (EIGHTH GRADE)
004 LKGE8-NO (4          ) 0010GEOGRAPHY: I DO NOT ENJOY TEACHING UNITED STATES HISTORY. (EIGHTH GRADE)
005 LKGE8-?  (M          ) 0001GEOGRAPHY: MISSING       (EIGHTH GRADE)
```

Table C-7

Conditioning Variables for the Long-Term Trend Reading Assessment

Conditioning Variable	Age Classes	Variable Name(s)	Variable Coding	Contrast Coding*
Overall	All			1
Gender	All	DSEX	Male	0
			Female	1
Size and Type of Community	All	STOC	Low Metro	00
			High Metro	10
			All others and Missing	01
Region	All	REGION	Northeast	000
			Southeast	100
			Central	010
			West	001
Parents' Education	All	PARED	Less than high school	0000
			High school graduate	1000
			Post-high school	0100
			College graduate	0010
			Missing and I Don't Know	0001
Items in the Home	All	B000901	None of the six items	00
		B000902	One of the six items	10
		B000903	Two of the six items	20
		B000904	Three of the six items	30
		B000905	Four of the six items	40
		B000906	Five of the six items	50
			Six of the six items	60
			Missing	01
Television Watching	All	B001801	None	00
			One hour or less	10
			Two hours	20
			Three hours	30
			Four hours	40
			Five hours	50
			Six or more hours	60
			Missing	01
Homework	All	B001701	Don't have any	00
			Don't do any	00
			Less than 1 hour	10
			1-2 hours	20
			More than 2 hours	30
			Missing	01

****Note*:** *Multicolumn entries without overbars indicate multiple contrasts. Barred columns are treated as one contrast.*

Table C-7 (continued)

Conditioning Variables for the Long-Term Trend Reading Assessment

Conditioning Variable	Age Classes	Variable Name(s)	Variable Coding	Contrast Coding*
Language Spoken at Home	All	B000401	English	00
			Spanish	10
			Other	10
			Missing	01
Pages Read	All	B001101	More than 20	10
			16-20	10
			11-15	10
			6-10	10
			5 or fewer	00
			Missing	01
Percent in School Lunch Program	All	PCLUNCH	0 percent	$\overline{000}$ 0
			1 percent	001 0
			2 percent	002 0
			.	.
			.	.
			99 percent	099 0
			100 percent	100 0
			Missing	000 1
Percent White	All	PCTWHT	0-49 Minority	100
			50-79 Integrated	010
			80-100 Predominantly White	001
			Missing	000
Courses Taken	9, 13	B001001	None of the seven	$\overline{00.0}$ 0
		B001002	One of the seven	01.0 0
		B002003	Two of the seven	02.0 0
		B002004	Three of the seven	03.0 0
		B002005	Four of the seven	04.0 0
		B002006	Five of the seven	05.0 0
		B002007	Six of the seven	06.0 0
			Seven of the seven	07.0 0
			Missing	00.0 1
Derived Race/Ethnicity	All	DRACE	White	000
			Black	100
			Hispanic	010
			Asian American	001
			American Indian	000
			Unclassified	000
			Missing	000

Note: *Multicolumn entries without overbars indicate multiple contrasts. Barred columns are treated as one contrast.*

Table C-7 (continued)

Conditioning Variables for the Long-Term Trend Reading Assessment

Conditioning Variable	Age Classes	Variable Name(s)	Variable Coding	Contrast Coding*
Age by Grade	All	MODGRD MODAGE	1 < modal age, modal grade	0000
			2 Modal age, < modal grade	
			3 Modal age, modal grade/missing	1000
			4 Modal age, > modal grade	0100
			5 > modal age, modal grade	
				0010
				0001

*Note: *Multicolumn entries without overbars indicate multiple contrasts. Barred columns are treated as one contrast.*

Table C-8

Conditioning Variables for the Long-Term Trend Mathematics Samples

Conditioning Variable	Age Classes	Variable Name(s)	Variable Coding	Contrast Coding*
Overall	All		---	1
Gender	All	DSEX	Male Female	0 1
Observed Race/Ethnicity	All	RACE	White Black Hispanic Asian American American Indian Other Missing	000 100 010 001 000 000 000
Size and Type of Community	All	STOC	1, 4-7 all except 2 and 3 2 Low Metro 3 High Metro	01 00 10
Region	All	REGION	Northeast Southeast Central West	000 100 010 001
Parents' Education	All	PARED	Less than high school High school graduate Post-high school College graduate Missing and I Don't Know	0000 1000 0100 0010 0001
Modal Grade	All	MODGRD	< modal grade = modal grade, missing > modal grade	10 00 01
Items in the Home (of newspaper, > 25 books, encyclopedia, magazines)	All	HOMEEN2	0 to 2 items 3 items 4 items	00 10 01

Note: *Multicolumn entries without overbars indicate multiple contrasts. Barred columns are treated as one contrast.*

Table C-8 (continued)

Conditioning Variables for the Long-Term Trend Mathematics Samples

Conditioning Variable	Age Classes	Variable Name(s)	Variable Coding	Contrast Coding*
Observed Race/Ethnicity by Gender ("White" includes American Indian and Other)	All	RACE DSEX	White, male	000
			Black, male	000
			Hispanic, male	000
			Asian American, male	000
			White, female	000
			Black, female	100
			Hispanic, female	010
			Asian American, female	001
Observed Race/Ethnicity by Parents' Education ("White" includes American Indian and Other)--coded differently for each age class	9	RACE PARED	White, < HS	0000 0000 0000
			White, HS graduate	0000 0000 0000
			White, post-HS	0000 0000 0000
			White, college grad.	0000 0000 0000
			White, missing	0000 0000 0000
			Black, < HS	0000 0000 0000
			Black, HS grad & post-HS	1000 0000 0000
			Black, college grad.	0010 0000 0000
			Black, missing	0001 0000 0000
			Hispanic, < HS	0000 0000 0000
			Hispanic, HS grad & post-HS	0000 1000 0000
			Hispanic, coll. grad.	0000 0010 0000
			Hispanic, missing	0000 0001 0000
			Asian Amer., < HS	0000 0000 0000
			Asian Amer., HS grad & post-HS	0000 0000 1000
			Asian Amer., coll. grad.	0000 0000 0010
			Asian Amer., missing	0000 0000 0001

__Note__: Multicolumn entries without overbars indicate multiple contrasts. Barred columns are treated as one contrast.

Table C-8 (continued)

Conditioning Variables for the Long-Term Trend Mathematics Samples

Conditioning Variable	Age Classes	Variable Name(s)	Variable Coding	Contrast Coding*
Observed Race/Ethnicity by Parents' Education ("White" includes American Indian and Other)--coded differently for each age class	13	RACE PARED	White, < HS	0000 0000 0000
			White, HS graduate	0000 0000 0000
			White, post-HS	0000 0000 0000
			White, college grad.	0000 0000 0000
			White, missing	0000 0000 0000
			Black, < HS	0000 0000 0000
			Black, HS graduate	1000 0000 0000
			Black, post-HS	0100 0000 0000
			Black, college grad.	0010 0000 0000
			Black, missing	0001 0000 0000
			Hispanic, < HS	0000 0000 0000
			Hispanic, HS grad.	0000 1000 0000
			Hispanic, post-HS	0000 0100 0000
			Hispanic, coll. grad.	0000 0010 0000
			Hispanic, missing	0000 0001 0000
			Asian Amer., < HS	0000 0000 0000
			Asian Amer., HS grad.	0000 0000 1000
			Asian Amer., post-HS	0000 0000 0100
			Asian Amer., coll. grad.	0000 0000 0010
			Asian Amer., missing	0000 0000 0001
Observed Race/Ethnicity by Parents' Education ("White" includes American Indian and Other)--coded differently for each age class	17	RACE PARED	White, < HS	0000 0000 0000
			White, HS graduate	0000 0000 0000
			White, post-HS	0000 0000 0000
			White, college grad.	0000 0000 0000
			White, missing	0000 0000 0000
			Black, < HS	0000 0000 0000
			Black, HS graduate	1000 0000 0000
			Black, post-HS	0100 0000 0000
			Black, college grad.	0010 0000 0000
			Black, missing	0001 0000 0000
			Hispanic, < HS	0000 0000 0000
			Hispanic, HS grad.	0000 1000 0000
			Hispanic, post-HS	0000 0100 0000
			Hispanic, coll. grad.	0000 0010 0000
			Hispanic, missing	0000 0001 0000
			Asian Amer., < HS	0000 0000 0000
			Asian Amer., HS grad.	0000 0000 1000
			Asian Amer., post-HS, coll. grad.	0000 0000 0100
			Asian Amer., missing	0000 0000 0001

Note: *Multicolumn entries without overbars indicate multiple contrasts. Barred columns are treated as one contrast.*

Table C-8 (continued)

Conditioning Variables for the Long-Term Trend Mathematics Samples

Conditioning Variable	Age Classes	Variable Name(s)	Variable Coding	Contrast Coding*
School Type	All	SCHTYPE	Public	0
			Private	1
			Catholic	1
			Bureau of Indian Affairs	1
			Department of Defense	1
Homework	13, 17	B003901	None assigned	100
			Didn't do	010
			½ hour or less	012
			1 hour	013
			2 hours	014
			More than 2 hours	000
			Missing	000
Language in the Home	All	LANGHOM	Never	00
			Sometimes	10
			Always	01
Observed Race/Ethnicity by Language in the Home--coded differently for age class 9	9	RACE LANGHOM	White, often	00 00 00
			White, sometimes	00 00 00
			White, never	00 00 00
			Black, often & sometimes	10 00 00
			Black, never	00 00 00
			Hispanic, often & sometimes	00 10 00
			Hispanic, never	00 00 00
			Asian Amer., often & sometimes	00 00 10
			Asian Amer., never	00 00 00
Observed Race/Ethnicity by Language in the Home	13, 17	RACE LANGHOM	White, often	00 00 00
			White, sometimes	00 00 00
			White, never	00 00 00
			Black, often	10 00 00
			Black, sometimes	01 00 00
			Black, never	00 00 00
			Hispanic, often	00 10 00
			Hispanic, sometimes	00 01 00
			Hispanic, never	00 00 00
			Asian Amer., often	00 00 10
			Asian Amer., sometimes	00 00 01
			Asian Amer., never	00 00 00

*__Note__: *Multicolumn entries without overbars indicate multiple contrasts. Barred columns are treated as one contrast.*

Table C-8 (continued)

Conditioning Variables for the Long-Term Trend Mathematics Samples

Conditioning Variable	Age Classes	Variable Name(s)	Variable Coding	Contrast Coding*
Highest Level Math Taken	17	NMATH	Pre-algebra Algebra Geometry Algebra 2 Calculus Something else	10000 01000 00100 00010 00001 00000
High School Program	17	B005001	General College preparatory Vocational, technical Missing	00 10 01 00
Derived Race/Ethnicity	All	DRACE	White Black Hispanic Asian American Other Missing	000 100 010 001 000 000

*__*Note*__*: Multicolumn entries without overbars indicate multiple contrasts. Barred columns are treated as one contrast.*

Table C-9

Conditioning Variables for the Long-Term Trend Science Assessment

Variable Name	Ages	Variable Coding	Contrast Coding
Overall	All		1
Gender (DSEX)	All	1 Male	0
		2 Female	1
Observed Race (RACE)	All	1 White	000
		2 Black	100
		3 Hispanic	010
		4 Asian American	001
		5 American Indian	000
		6 Other, blank, missing	000
Size and Type of Community (STOC)	All (92 only)	1, 4-7, blank, missing	00
		2 Low Metro	10
		3 High Metro	01
Type of Location (TOL8)	ALL (94 only)	1 Big City	00000000
		2 Medium City	10000000
		3 Fringe of Big City	01000000
		4 Fringe of Medium City	00100000
		5 Large Town	00010000
		6 Small Place	00001000
		7 Rural - MSA	00000100
		8 Rural - Non MSA	00000010
		Missing	00000001
REGION	All	1 Northeast, missing	000
		2 Southeast	100
		3 Central	010
		4 West	001
		Missing	000
Parents' Education (PARED)	All	1 < High School	0000
		2 = High School	1000
		3 > High School	0100
		4 Graduated College	0010
		5 Unknown, missing	0001
Modal Grade (MODGRD)	All	1 < Modal Grade	10
		2 = Modal Grade, missing	00
		3 > Modal Grade	01

Table C-9 (continued)

Conditioning Variables for the Long-Term Trend Science Assessment

Variable Name	Ages	Variable Coding	Contrast Coding
Observed Race x Gender (RACE x DSEX)	All	1 White - Male	000
		2 Black - Male	000
		3 Hispanic - Male	000
		4 Asian American - Male	000
		5 White - Female	000
		6 Black - Female	100
		7 Hispanic - Female	010
		8 Asian American - Female	001
		Other combinations, missing	000
Observed Race x Parents' Education (RACE x PARED)	All	1 White - < High School	0000 0000 0000
		2 White - = High School	0000 0000 0000
		3 White - > High School	0000 0000 0000
		4 White - Graduated College	0000 0000 0000
		5 White - Missing, Unknown	0000 0000 0000
		6 Black - < High School	0000 0000 0000
		7 Black - = High School	1000 0000 0000
		8 Black - > High School	0100 0000 0000
		9 Black - Graduated College	0010 0000 0000
		10 Black - Missing, Unknown	0001 0000 0000
		11 Hispanic - < High School	0000 0000 0000
		12 Hispanic - = High School	0000 1000 0000
		13 Hispanic - > High School	0000 0100 0000
		14 Hispanic - Graduated College	0000 0010 0000
		15 Hispanic - Missing, Unknown	0000 0001 0000
		16 Asian Amer. - < High School	0000 0000 0000
		17 Asian Amer. - = High School	0000 0000 1000
		18 Asian Amer. - > High School	0000 0000 0100
		19 Asian Amer. - Graduated College	0000 0000 0010
		20 Asian Amer. - Missing, Unknown	0000 0000 0001
School Type (SCHTYPE)	All	1 Public, missing	0
		2 Private	1
		3 Catholic	1
		4 Bureau of Indian Affairs	1
		5 Department of Defense	1
Items in the Home (HOMEEN2)	All	1 0-2 Items, missing	00
		2 3 Items	10
		3 4 Items	01

Table C-9 (continued)

Conditioning Variables for the Long-Term Trend Science Assessment

Variable Name	Ages	Variable Coding	Contrast Coding
Derived Race (DRACE)	All	1 White	000
		2 Black	100
		3 Hispanic	010
		4 Asian American	001
		5 American Indian	000
		6 Other, missing	000
Language in the Home (LANGHOM)	All	1 Never, missing	00
		2 Sometimes	10
		3 Always	01
Observed Race x Language in the Home (RACE x LANGHOM)	All	1 White - Often	00 00 00
		2 White - Sometimes	00 00 00
		3 White - Never	00 00 00
		4 Black - Often	10 00 00
		5 Black - Sometimes	01 00 00
		6 Black - Never	00 00 00
		7 Hispanic - Often	00 10 00
		8 Hispanic - Somestimes	00 01 00
		9 Hispanic - Never	00 00 00
		10 Asian Amer. - Often	00 00 10
		11 Asian Amer. - Sometimes	00 00 01
		12 Asian Amer. - Never	00 00 00
		One or both missing	00 00 00
Homework (B003901)	13,17	1 None Assigned	100
		2 Didn't Do	010
		3 1/2 Hour or Less	012
		4 One Hour	013
		5 Two Hours	014
		6 More Than Two Hours	000
		missing	000
NSCI (B005308-B005311)	17	1 General Science	1000
		2 Biology	0100
		3 Chemistry	0010
		4 Physics	0001
		5 Nothing, Something Else	0000
		missing	0000
High School Program (B005001)	17	1 General	00
		2 College Preparatory	10
		3 Vocational, Technical	01
		missing	00

Table C-10

Conditioning Variables for the Long-Term Trend Writing Assessment

Writing Trend Conditioning Variables				
	Variable	**Number of Contrasts**	**Contrasts**	**Contrasts Codes**
Intercept		1	Overall	1
Gender	DSEX	1	Male	0
			Female	1
Race/Ethnicity	Ethnic(84)	2	Black	00
	DRACE		Hispanic	01
	(88,90,92,94)		White+others	10
Type of Community/ Location	STOC	2	STOC:	STOC:
	(84,88,90,92)		Low Metro	00
			Everyone else	01
			Hi Metro	10
	TOL8(94)	7	TOL8:	
			Large City & miss	0000000
			Medium City	1000000
			Urb. Fring of LC	0100000
			Urb. Fring of MC	0010000
			Big Town	0001000
			Small Town	0000100
			Rural-MSA	0000010
			Rural-non-MSA	0000001
Region	REGION	3	Northeast	000
			Southeast	100
			Central	010
			West	001
Parent's Education	PARED	4	Less than H.S.	0000
			High School	1000
			Greater HS	0100
			College	0010
			I don't know	0001
Items in the home	B000901	2	Articles in home:	
	B000902		0-3 or missing	
	B000903		4	00
	B000904		5	10
	B000905			01
Modal age	DAGE	2	Modal Age	00
			< Modal age	10
			> Modal age	01

Table C-10 (continued)

Conditioning Variables for the Long-Term Trend Writing Assessment

Writing Trend Conditioning Variables				
	Variable	**Number of Contrasts**	**Contrasts**	**Contrasts Codes**
Homework	B001701	5	None assigned None done < 1 hr 1-2 hrs 2+ hrs Missing HW	00000 10000 01000 00100 00010 00001
School type	SCHTYPE	2	Public School Non-Public (Priv,Cath,Dod BIA) and Miss	0 1
TV watching, Linear	B001801	1	zero one hour two hours Three hours or Missing four hours five hours six hours	0 1 2 3 4 5 6
TV watching Quadratic	B001801	2	zero one hour two hours Three hours or Missing four hours five hours six hours	00 01 04 09 16 25 36
Mother Work Outside the Home	B000801	1	Mother does not work or missing Mother works	0 1
Language Minority (self speaking at home	B000301	2	Not Language Minority Language Minority Missing	00 10 01

Table C-10 (continued)

Conditioning Variables for the Long-Term Trend Writing Assessment

Writing Trend Conditioning Variables				
	Variable	**Number of Contrasts**	**Contrasts**	**Contrasts Codes**
Percent in Lunch Program	C001301(84) PCLUNCH (88,90) C0320101(92,94)	7	none in lunch prog 1-5% 6-10% 11-25% 26-50% 51-75% 76-90% over 90%	0000000 1000000 0100000 0010000 0001000 0000100 0000010 0000001
Minority School	PCTWHT PCTWHTQ	2	Minority (0-49.9%) Intergrated (50- 79.9%) White or Missing (80-100%)	00 10 01
Grades in School	B001901	1 Var	Grade (Missing included at 2.0)	0.0-4.0
Pages Read for HW	B001101	3	No pages read 11+ pages read 6-10 pages read 5 or less pages read	000 100 010 001
Number of reports	B001201-7 and B001208	1	Number of reports, essays etc.	0-7

Table C-11

Proportion of Variance of the Conditioning Variable Contrasts Accounted for by the Principal Components Used in the Conditioning Model for Reading Main Conditioning Variables, Age 9/Grade 4

Contrast	Proportion of Variance	Contrast	Proportion of Variance	Contrast	Proportion of Variance
FEMALE	0.925707	HW-NO	0.989857	G/S 23	0.898663
BLACK	0.940444	HW-YES	0.990370	R/T 22	0.899784
HISPANIC	0.968477	HWLIN-0	0.984957	R/T 23	0.902370
ASIAN	0.975347	HWQUAD-0	0.983435	R/T 24	0.881287
PAC ISLD	0.988082	HITEM=3	0.858982	R/T 25	0.912624
MEXICAN	0.963736	HITEM=4	0.812504	R/T 32	0.890183
PUER RIC	0.970751	MOMHOM-N	0.953431	R/T 33	0.889589
CUBN,OTH	0.978096	MOMHOM-?	0.788223	R/T 34	0.872553
HISP-?	0.960605	DADHOM-N	0.931288	R/T 35	0.887940
NON MSA	0.863202	DADHOM-?	0.714905	R/T 42	0.958209
MID CTY5	0.937670	MISS-2<	0.988519	R/T 43	0.960167
FR/BTWN5	0.938775	USA 3-5	0.977317	R/T 44	0.963760
SML TWN5	0.938518	USA <3	0.975098	R/T 45	0.965497
RURAL5	0.950218	USA-?	0.739579	R/T 52	0.977929
URBAN FR	0.884268	STGRD1-2	0.890737	R/T 53	0.973307
MED CITY	0.833305	STGRD3>	0.897091	R/T 54	0.968598
SM PLACE	0.902924	PRESCH-N	0.971256	R/T 55	0.971651
HS GRAD	0.950195	SCHCHG-1	0.883960	R/P 22	0.909081
POST HS	0.932234	SCHCHG-2	0.903256	R/P 23	0.899331
COL GRAD	0.929429	SCHCHG-3	0.841302	R/P 24	0.913507
PARED-?	0.936772	DIS@HOM2	0.899320	R/P 25	0.914537
S EAST	0.898570	DIS@HOM3	0.933883	R/P 32	0.901818
CENTRAL	0.857113	DIS@HOM4	0.839289	R/P 33	0.906250
WEST	0.886716	PGS>5	0.813264	R/P 34	0.909226
PRIVATE	0.924010	PGS>10	0.817465	R/P 35	0.905588
CATHOLIC	0.916306	COMP-WK	0.941506	R/P 42	0.981855
IEP-NO	0.973795	COMP-MO	0.963296	R/P 43	0.973368
LEP-NO	0.918151	COMP-NEV	0.931707	R/P 44	0.975298
CHAP1-N	0.900958	COMP-?	0.686891	R/P 45	0.975382
INTEGRAT	0.869162	G/R 22	0.919363	R/P 52	0.986931
MINORITY	0.831395	G/R 23	0.916281	R/P 53	0.983883
NO LUNCH	0.826878	G/R 24	0.975892	R/P 54	0.986325
LUNCH-?	0.830445	G/R 25	0.987520	R/P 55	0.986780
=MA/<MG	0.882317	G/T 22	0.877126	R/S 22	0.969271
=MA/=MG	0.962432	G/T 23	0.827592	R/S 23	0.974890
=MA/>MG	0.992707	G/T 24	0.729063	R/S 32	0.964997
>MA/=MG	0.968290	G/T 25	0.739566	R/S 33	0.947360
HL-SOME	0.942233	G/P 22	0.941166	R/S 42	0.979138
HL-ALWAY	0.959744	G/P 23	0.896848	R/S 43	0.982155
HL-?	0.787530	G/P 24	0.853280	R/S 52	0.985796
TV-LIN	0.985159	G/P 25	0.860958	R/S 53	0.980516
TV-QUAD	0.985013	G/S 22	0.922181	T/P 22	0.816668

Table C-11 (continued)

Proportion of Variance of the Conditioning Variable Contrasts Accounted for by the Principal Components Used in the Conditioning Model for Reading Main Conditioning Variables, Age 9/Grade 4

Contrast	Proportion of Variance	Contrast	Proportion of Variance	Contrast	Proportion of Variance
T/P 23	0.858029	RD4FUN-?	0.704829	TK2LIB-3	0.937926
T/P 24	0.803055	FAMRED-2	0.905949	TK2LIB-4	0.940071
T/P 25	0.819084	FAMRED-3	0.920054	TK2LIB-?	0.807041
T/P 32	0.776803	FAMRED-4	0.885275	PFOLIO-N	0.949504
T/P 33	0.829405	FAMRED-?	0.776441	PFOLIO-?	0.719388
T/P 34	0.800320	USELIB-2	0.909888	#QUEST+2	0.871530
T/P 35	0.802270	USELIB-3	0.928996	#QUEST+3	0.858844
T/P 42	0.834125	USELIB-4	0.938013	#QUEST+4	0.912630
T/P 43	0.887889	USELIB-?	0.778238	#QUEST+?	0.889205
T/P 44	0.794982	VOCAB-2	0.881024	TESTDIF2	0.921518
T/P 45	0.770371	VOCAB-3	0.916387	TESTDIF3	0.906347
T/P 52	0.808895	VOCAB-4	0.879758	TESTDIF4	0.915550
T/P 53	0.879782	VOCAB-?	0.804105	TESTDIF?	0.933663
T/P 54	0.807940	TALKRD-2	0.926773	TESTEFF2	0.881544
T/P 55	0.780056	TALKRD-3	0.935181	TESTEFF3	0.850143
T/S 22	0.919122	TALKRD-4	0.910212	TESTEFF4	0.917166
T/S 23	0.908739	TALKRD-?	0.791504	TESTEFF?	0.866815
T/S 32	0.923196	WBK/WS-2	0.913055	TESTIMP2	0.927168
T/S 33	0.903300	WBK/WS-3	0.923436	TESTIMP3	0.946176
T/S 42	0.934263	WBK/WS-4	0.921429	TESTIMP4	0.953298
T/S 43	0.927077	WBK/WS-?	0.792329	TESTIMP?	0.930972
T/S 52	0.951463	WRTRED-2	0.916018	LNGANSW2	0.856334
T/S 53	0.941835	WRTRED-3	0.919121	LNGANSW3	0.877543
P/S 22	0.965610	WRTRED-4	0.919896	LNGANSW4	0.866921
P/S 23	0.947750	WRTRED-?	0.797807	LNGANSW?	0.936771
P/S 32	0.959227	RDPRJ-S2	0.934545	SNRM-LIN	0.803585
P/S 33	0.942198	RDPRJ-S3	0.930419	DEPTLIZD	0.867085
P/S 42	0.945685	RDPRJ-S4	0.929365	REGROUPD	0.839255
P/S 43	0.926091	RDPRJ-S?	0.712197	SCH4ORG?	0.838592
P/S 52	0.947819	ALOUD-S2	0.913447	CLAABL-N	0.826823
P/S 53	0.931281	ALOUD-S3	0.920746	CLAABL-?	0.883128
NBOOK-12	0.956602	ALOUD-S4	0.909066	RD PRI-N	0.864144
NBOOK-34	0.963321	ALOUD-S?	0.744083	RD PRI-?	0.930351
NBOOK-5+	0.950080	SILNT-S2	0.929079	%SUBLUN2	0.880605
NBOOK-?	0.853271	SILNT-S3	0.941801	%SUBLUN3	0.919789
GOOD RDR	0.860588	SILNT-S4	0.932255	%SUBLUN4	0.892617
AVG RDR	0.861515	SILNT-S?	0.680063	%SUBLUN5	0.874687
POOR RDR	0.935414	OWNBK-S2	0.890039	%SUBLUN?	0.904112
RDR-MISS	0.873855	OWNBK-S3	0.891576	%REMRED2	0.891034
RD4FUN-2	0.887341	OWNBK-S4	0.891536	%REMRED3	0.892055
RD4FUN-3	0.904736	OWNBK-S?	0.776048	%REMRED4	0.845198
RD4FUN-4	0.871175	TK2LIB-2	0.906528	%REMRED5	0.831543

Table C-11 (continued)

Proportion of Variance of the Conditioning Variable Contrasts Accounted for by the Principal Components Used in the Conditioning Model for Reading Main Conditioning Variables, Age 9/Grade 4

Contrast	Proportion of Variance	Contrast	Proportion of Variance	Contrast	Proportion of Variance
%REMRED?	0.893915	TMCH-COM	0.993098	TDEG-?	0.972416
%ENREOY2	0.884100	T_FEMALE	0.820295	UGRENG-?	0.865807
%ENREOY3	0.897341	T_SEX-?	0.981626	UGRRED-?	0.847997
%ENREOY4	0.882357	T_YREXP2	0.921128	GRDENG-?	0.885304
%ENREOY?	0.920674	T_YREXP3	0.910245	GRDRED-?	0.851862
PARAID-O	0.880765	T_YREXP4	0.930107	READDEV2	0.922295
PARAID-N	0.893730	T_YREXP5	0.946359	READDEV3	0.929516
PARAID-?	0.978688	T_YREXP?	0.983411	READDEV4	0.929900
PARHW-N?	0.853677	T_BLACK	0.821855	READDEV5	0.927907
PARHW-N?	0.922685	T_HISP	0.907762	READDEV5	0.976188
S/PHW-N?	0.849648	T_ASIAN	0.922536	PRPER-<1	0.912015
S/PHW-N?	0.825103	T_PAC IS	0.855421	PRPER-12	0.915907
VOLPG-N?	0.881626	T_RACE-?	0.961165	PRPER->2	0.921574
VOLPG-N?	0.815107	CERTG-N	0.886028	PRPER-?	0.963691
SCHCH1-N	0.828842	CERTG-?	0.961575	CSPECS-N	0.746273
SCHCH1-?	0.964249	CERTR-N	0.883596	CSPECS-?	0.939094
%CHAP1-2	0.871525	CERTR-NS	0.840139	CLASIZ-2	0.861864
%CHAP1-3	0.883095	CERTR-?	0.908862	CLASIZ-3	0.882162
%CHAP1-4	0.872850	CERTL-N	0.912746	CLASIZ-4	0.864424
%CHAP1-5	0.839106	CERTL-NS	0.888155	CLASIZ-5	0.904019
%CHAP1-?	0.794374	CERTL-?	0.917385	CLASIZ-?	0.990065
SCH%RURL	0.849219	RESOURC2	0.909317	T_ABIL-N	0.824050
RUR%-LIN	0.833293	RESOURC3	0.923936	T_ABIL-?	0.988419
SCH%CITY	0.845520	RESOURC4	0.958483	ABILRED2	0.923965
CITY%-LN	0.889416	RESOURC?	0.977820	ABILRED3	0.907574
PAR%PROF	0.932039	UGR ED-Y	0.799656	ABILRED4	0.943970
PROF%-LN	0.838742	GRA ED-Y	0.833415	ABILRED?	0.977145
PAR%BCOL	0.940609	NOGRAD-Y	0.860866	INSTIME2	0.921612
BCOL%-LN	0.833693	T4REDYR2	0.920304	INSTIME3	0.944921
PAR%FARM	0.870251	T4REDYR3	0.908436	INSTIME4	0.926513
FARM%-LN	0.839498	T4REDYR4	0.928472	INSTIME?	0.983312
PAR%IRRE	0.914334	T4REDYR5	0.946441	FLEX GRP	0.823984
IRRE%-LN	0.802939	T4REDYR?	0.980263	2 GROUPS	0.866268
PAR%WELF	0.929665	CERT-TMP	0.934858	3 GROUPS	0.858408
WELF%-LN	0.861288	CERT-REG	0.948042	4 GROUPS	0.902751
LSTAF-NO	0.895306	CERT-HGH	0.953207	5+GROUPS	0.907832
LSTAF-PT	0.958098	CERT-?	0.960546	INDIVLZD	0.843579
LSTAF-FT	0.960129	TDEG-ASC	0.961422	GROUPS-?	0.976673
CMPCL-N?	0.840618	TDEG-BAC	0.936737	TRADE	0.870715
CMPLB-N?	0.824930	TDEG-MAS	0.940835	BAS/TRAD	0.823012
CMPLB-N?	0.864775	TDEG-EDS	0.927094	OTHER RM	0.879333
TMCH-PAR	0.893581	TDEG-PRO	0.976748	RDMATS-?	0.925340

Table C-11 (continued)

Proportion of Variance of the Conditioning Variable Contrasts Accounted for by the Principal Components Used in the Conditioning Model for Reading Main Conditioning Variables, Age 9/Grade 4

Contrast	Proportion of Variance	Contrast	Proportion of Variance	Contrast	Proportion of Variance
CHIMAG-2	0.945656	OWNBK-T3	0.859006	CLSLIB-3	0.920999
CHIMAG-3	0.949745	OWNBK-T4	0.892341	CLSLIB-4	0.834620
CHIMAG-4	0.933693	OWNBK-T?	0.991971	CLSLIB-5	0.865637
CHIMAG-?	0.978583	PRJCT-T2	0.914697	CLSLIB-?	0.992053
RDKITS-1	0.902222	PRJCT-T3	0.940036	CLALIB-2	0.894031
RDKITS-2	0.923596	PRJCT-T4	0.839021	CLALIB-3	0.896787
RDKITS-3	0.917627	INTRP-T2	0.874561	CLALIB-4	0.885843
RDKITS-?	0.966317	INTRP-T3	0.903045	CLALIB-5	0.870351
SOFTWR-2	0.922557	INTRP-T4	0.861242	CLALIB-5	0.970747
SOFTWR-3	0.916982	INTRP-T?	0.981849	COMP-DIF	0.885807
SOFTWR-4	0.927168	EXPLA-T2	0.850604	COMP-AVL	0.880050
SOFTWR-?	0.978683	EXPLA-T3	0.866013	COMP-?	0.975712
VARBKS-2	0.840442	EXPLA-T4	0.875867		
VARBKS-3	0.864488	EXPLA-T?	0.992001		
VARBKS-4	0.833996	QUIZS-T2	0.932978		
VARBKS-?	0.989874	QUIZS-T3	0.946134		
OTHMAT-2	0.845529	QUIZS-T4	0.850842		
OTHMAT-3	0.871562	QUIZA-T?	0.992711		
OTHMAT-4	0.863954	MOVIE-T2	0.949565		
OTHMAT-?	0.990592	MOVIE-T3	0.968929		
VOCAB-T2	0.797891	MOVIE-T4	0.923741		
VOCAB-T3	0.852970	MCTEST-2	0.876984		
VOCAB-T4	0.825991	MCTEST-3	0.918790		
VOCAB-T?	0.994164	MCTEST-4	0.861967		
ALOUD-T2	0.831957	MCTEST-?	0.978188		
ALOUD-T3	0.886286	SATEST-2	0.842366		
ALOUD-T4	0.876477	SATEST-3	0.897165		
TLKRD-T2	0.859493	SATEST-4	0.851853		
TLKRD-T3	0.889382	SATEST-?	0.984811		
TLKRD-T4	0.830299	WRITST-2	0.841600		
WRTRD-T2	0.846517	WRITST-3	0.858246		
WRTRD-T3	0.895640	WRITST-4	0.843669		
WRTRD-T4	0.823382	WRITST-?	0.978531		
WRTRD-T?	0.993035	PRJTST-2	0.900467		
WB/WS-T2	0.845151	PRJTST-3	0.922259		
WB/WS-T3	0.879112	PRJTST-4	0.828960		
WB/WS-T4	0.873668	PRJTST-?	0.987599		
SILNT-T2	0.818774	PRTTST-2	0.907816		
SILNT-T3	0.905303	PRTTST-3	0.880670		
SILNT-T4	0.955160	PRTTST-4	0.881917		
SILNT-T?	0.969065	PRTTST-?	0.979132		
OWNBK-T2	0.848959	CLSLIB-2	0.939617		

Table C-12

Proportion of Variance of the Conditioning Variable Contrasts Accounted for by the Principal Components Used in the Conditioning Model for Reading Main Conditioning Variables, Age 13/Grade 8

Contrast	Proportion of Variance	Contrast	Proportion of Variance	Contrast	Proportion of Variance
FEMALE	0.923029	HW-NO	0.975695	G/S 23	0.943016
BLACK	0.923760	HW-YES	0.976868	R/T 25	0.898101
HISPANIC	0.971534	HWLIN-0	0.980037	R/T 31	0.882716
ASIAN	0.968087	HWQUAD-0	0.978694	R/T 32	0.884500
PAC ISLD	0.968000	HITEM=3	0.882134	R/T 33	0.900235
MEXICAN	0.955524	HITEM=4	0.846224	R/T 34	0.889987
PUER RIC	0.959221	MOMHOM-N	0.986350	R/T 35	0.896666
CUBN,OTH	0.976382	MOMHOM-?	0.884483	R/T 41	0.890105
HISP-?	0.984748	DADHOM-N	0.963363	R/T 42	0.901502
NON MSA	0.857439	DADHOM-?	0.900600	R/T 43	0.957868
MID CTY5	0.923970	MISS-2<	0.984879	R/T 44	0.955803
FR/BTWN5	0.927039	USA 3-5	0.965674	R/T 45	0.961614
SML TWN5	0.934399	USA <3	0.962805	R/T 51	0.971644
RURAL5	0.945599	USA-?	0.823240	R/T 52	0.972940
URBAN FR	0.859548	STGRD1-2	0.954732	R/T 53	0.970104
MED CITY	0.837783	STGRD3-5	0.942703	R/T 54	0.961403
SM PLACE	0.889482	STGRD6>	0.912122	R/T 55	0.975837
HS GRAD	0.956768	SCHCHG-1	0.940970	R/P 25	0.910028
POST HS	0.947736	SCHCHG-2	0.938478	R/P 31	0.914335
COL GRAD	0.937187	SCHCHG-3	0.738295	R/P 32	0.914324
PARED-?	0.959069	DIS@HOM2	0.923974	R/P 33	0.864237
S EAST	0.889366	DIS@HOM3	0.936882	R/P 34	0.907824
CENTRAL	0.889377	DIS@HOM4	0.825385	R/P 35	0.913140
WEST	0.865575	PGS>5	0.795626	R/P 41	0.922807
PRIVATE	0.934604	PGS>10	0.804433	R/P 42	0.876326
CATHOLIC	0.906925	COMP-WK	0.959685	R/P 43	0.985957
IEP-NO	0.974905	COMP-MO	0.949271	R/P 44	0.981514
LEP-NO	0.881836	COMP-NEV	0.938234	R/P 45	0.977946
CHAP1-N	0.872352	COMP-?	0.817228	R/P 51	0.962999
INTEGRAT	0.891808	G/R 22	0.910441	R/P 52	0.986195
MINORITY	0.828774	G/R 23	0.911370	R/P 53	0.988716
NO LUNCH	0.773265	G/R 24	0.973356	R/P 54	0.985942
LUNCH-?	0.869466	G/R 25	0.979936	R/P 55	0.979845
=MA/<MG	0.800935	G/T 22	0.852894	R/S 32	0.971119
=MA/=MG	0.942686	G/T 23	0.871642	R/S 33	0.970852
=MA/>MG	0.970862	G/T 24	0.712349	R/S 41	0.976346
>MA/=MG	0.945148	G/T 25	0.738634	R/S 42	0.947132
HL-SOME	0.863552	G/P 22	0.929049	R/S 43	0.982209
HL-ALWAY	0.795811	G/P 23	0.947156	R/S 51	0.977373
HL-?	0.830948	G/P 24	0.841356	R/S 52	0.988204
TV-LIN	0.985999	G/P 25	0.792744	R/S 53	0.972702
TV-QUAD	0.985677	G/S 22	0.958026	T/P 25	0.794251

Table C-12 (continued)

Proportion of Variance of the Conditioning Variable Contrasts Accounted for by the Principal Components Used in the Conditioning Model for Reading Main Conditioning Variables, Age 13/Grade 8

Contrast	Proportion of Variance	Contrast	Proportion of Variance	Contrast	Proportion of Variance
T/P 31	0.817048	RD4FUN-?	0.783814	ALOUD-S3	0.900077
T/P 32	0.736476	FAMRED-2	0.945769	ALOUD-S4	0.906567
T/P 33	0.847137	FAMRED-3	0.951671	ALOUD-S?	0.655516
T/P 34	0.745214	FAMRED-4	0.940412	SILNT-S2	0.909371
T/P 35	0.767101	FAMRED-?	0.774252	SILNT-S3	0.912068
T/P 41	0.712177	NOVEL-2	0.951360	SILNT-S4	0.928784
T/P 42	0.827660	NOVEL-3	0.946718	SILNT-S?	0.684191
T/P 43	0.821789	NOVEL-4	0.942295	OWNBK-S2	0.929455
T/P 44	0.836644	NOVEL-?	0.668399	OWNBK-S3	0.933383
T/P 45	0.814339	USELIB-2	0.955504	OWNBK-S4	0.918511
T/P 51	0.888475	USELIB-3	0.965932	OWNBK-S?	0.794986
T/P 52	0.782582	USELIB-4	0.943149	EXPLA-2	0.934227
T/P 53	0.814832	USELIB-?	0.869109	EXPLA-3	0.926618
T/P 54	0.793880	RDNEWS-2	0.901595	EXPLA-4	0.921522
T/P 55	0.863922	RDNEWS-3	0.918731	EXPLA-?	0.831286
T/S 32	0.915087	RDNEWS-4	0.908182	DISCU-2	0.938022
T/S 33	0.915932	RDNEWS-?	0.901441	DISCU-3	0.943521
T/S 41	0.920947	RDMAGZ-2	0.873682	DISCU-4	0.935194
T/S 42	0.903327	RDMAGZ-3	0.878675	DISCU-?	0.853983
T/S 43	0.941724	RDMAGZ-4	0.914430	AC2LIB-N	0.975313
T/S 51	0.929878	RDMAGZ-?	0.657614	AC2LIB-?	0.767389
T/S 52	0.942830	VOCAB-2	0.910060	LIB4RS-2	0.925494
T/S 53	0.938227	VOCAB-3	0.925770	LIB4RS-3	0.946218
P/S 32	0.972958	VOCAB-4	0.906744	LIB4RS-4	0.964531
P/S 33	0.953270	VOCAB-?	0.815649	LIB4RS-5	0.886947
P/S 41	0.963867	TALKRD-2	0.958244	LIB4RS-?	0.857091
P/S 42	0.945677	TALKRD-3	0.961268	LIB2BW-2	0.923590
P/S 43	0.955981	TALKRD-4	0.952475	LIB2BW-3	0.953526
P/S 51	0.934578	TALKRD-?	0.785864	LIB2BW-4	0.964809
P/S 52	0.976867	WBK/WS-2	0.906454	LIB2BW-5	0.889795
P/S 53	0.957014	WBK/WS-3	0.911827	LIB2BW-?	0.870313
NBOOK-12	0.910867	WBK/WS-4	0.899968	LIB4U-2	0.947402
NBOOK-34	0.926659	WBK/WS-?	0.714868	LIB4U-3	0.947712
NBOOK-5+	0.936535	WRTRED-2	0.930098	LIB4U-4	0.971038
NBOOK-?	0.868545	WRTRED-3	0.918677	LIB4U-5	0.911776
GOOD RDR	0.906303	WRTRED-4	0.935711	LIB4U-?	0.880551
AVG RDR	0.886993	WRTRED-?	0.685709	LIB4NJ-2	0.886007
POOR RDR	0.944202	RDPRJ-S2	0.962029	LIB4NJ-3	0.936756
RDR-MISS	0.753937	RDPRJ-S3	0.959238	LIB4NJ-4	0.973705
RD4FUN-2	0.938400	RDPRJ-S4	0.959481	LIB4NJ-5	0.859770
RD4FUN-3	0.933415	RDPRJ-S?	0.749335	LIB4NJ-?	0.886990
RD4FUN-4	0.917210	ALOUD-S2	0.903582	LIB4QI-2	0.931796

Table C-12 (continued)

Proportion of Variance of the Conditioning Variable Contrasts Accounted for by the Principal Components Used in the Conditioning Model for Reading Main Conditioning Variables, Age 13/Grade 8

Contrast	Proportion of Variance	Contrast	Proportion of Variance	Contrast	Proportion of Variance
LIB4QI-3	0.931831	%REMRED5	0.815961	CMPCL-N?	0.849162
LIB4QI-4	0.945053	%REMRED?	0.934085	CMPLB-N?	0.754063
LIB4QI-5	0.882260	%ENREOY2	0.872535	CMPLB-N?	0.841312
LIB4QI-?	0.896635	%ENREOY3	0.907089	TMCH-PAR	0.883983
PFOLIO-N	0.972638	%ENREOY4	0.910312	TMCH-COM	0.978847
PFOLIO-?	0.782274	%ENREOY?	0.917644	T_FEMALE	0.794995
#QUEST+2	0.893282	PARAID-O	0.877892	T_SEX-?	0.983035
#QUEST+3	0.884923	PARAID-N	0.905602	T_YREXP2	0.867232
#QUEST+4	0.932089	PARAID-?	0.984093	T_YREXP3	0.865803
#QUEST+?	0.945477	PARHW-N?	0.864624	T_YREXP4	0.909609
TESTDIF2	0.970324	PARHW-N?	0.819011	T_YREXP5	0.904769
TESTDIF3	0.960463	S/PHW-N?	0.907220	T_YREXP?	0.986117
TESTDIF4	0.959060	S/PHW-N?	0.918204	T_BLACK	0.845129
TESTDIF?	0.968003	VOLPG-N?	0.877146	T_HISP	0.887667
TESTEFF2	0.930380	VOLPG-N?	0.812284	T_ASIAN	0.883269
TESTEFF3	0.918544	SCHCH1-N	0.825729	T_AM IND	0.980443
TESTEFF4	0.942236	SCHCH1-?	0.956235	T_RACE-?	0.971322
TESTEFF?	0.951247	%CHAP1-2	0.875716	CERTG-N	0.807405
TESTIMP2	0.923706	%CHAP1-3	0.872001	CERTG-NS	0.958141
TESTIMP3	0.906879	%CHAP1-4	0.829223	CERTG-?	0.838638
TESTIMP4	0.910109	%CHAP1-5	0.855466	CERTR-N	0.845360
TESTIMP?	0.967384	%CHAP1-?	0.750356	CERTR-NS	0.835663
LNGANSW2	0.850308	SCH%RURL	0.802093	CERTR-?	0.872262
LNGANSW3	0.865303	RUR%-LIN	0.868842	CERTL-N	0.842684
LNGANSW4	0.917511	SCH%CITY	0.827590	CERTL-NS	0.843767
LNGANSW?	0.968041	CITY%-LN	0.864498	CERTL-?	0.861018
SNRM-LIN	0.808255	PAR%PROF	0.961823	CERTE-N	0.876950
SEMIDEPT	0.918770	PROF%-LN	0.849917	CERTE-NS	0.856870
DEPTLIZD	0.934735	PAR%BCOL	0.960000	CERTE-?	0.850301
SCH8ORG?	0.895551	BCOL%-LN	0.845786	RESOURC2	0.871316
ENGABL-N	0.807632	PAR%FARM	0.858293	RESOURC3	0.898699
ENGABL-?	0.954674	FARM%-LN	0.819698	RESOURC4	0.902948
RD PRI-N	0.809402	PAR%IRRE	0.940796	RESOURC?	0.986981
RD PRI-?	0.904156	IRRE%-LN	0.784950	UGR ED-Y	0.757307
%SUBLUN2	0.876542	PAR%WELF	0.950555	GRA ED-Y	0.803399
%SUBLUN3	0.869826	WELF%-LN	0.787409	NOGRAD-Y	0.835559
%SUBLUN4	0.904071	LSTAF-NO	0.892048	T8REDYR2	0.876160
%SUBLUN5	0.873188	LSTAF-PT	0.940978	T8REDYR3	0.866000
%SUBLUN?	0.922027	LSTAF-FT	0.955399	T8REDYR4	0.867088
%REMRED2	0.878681	LSTAF-?	0.980840	T8REDYR5	0.899404
%REMRED3	0.880667	TCHSUB=1	0.778694	T8REDYR?	0.945810
%REMRED4	0.836042	TCHSUB-?	0.981329	CERT-TMP	0.883379

Table C-12 (continued)

Proportion of Variance of the Conditioning Variable Contrasts Accounted for by the Principal Components Used in the Conditioning Model for Reading Main Conditioning Variables, Age 13/Grade 8

Contrast	Proportion of Variance	Contrast	Proportion of Variance	Contrast	Proportion of Variance
CERT-REG	0.926610	5+GROUPS	0.945123	WB/WS-T?	0.985402
CERT-HGH	0.949680	INDIVLZD	0.832963	SILNT-T2	0.822277
CERT-?	0.949701	GROUPS-?	0.937301	SILNT-T3	0.881450
TDEG-ASC	0.947320	TRADE	0.873832	SILNT-T4	0.894185
TDEG-BAC	0.925052	BAS/TRAD	0.858208	SILNT-T?	0.976750
TDEG-MAS	0.928104	OTHER RM	0.891629	OWNBK-T2	0.860465
TDEG-EDS	0.929013	RDMATS-?	0.897184	OWNBK-T3	0.908278
TDEG-DOC	0.937988	RDKITS-1	0.924425	OWNBK-T4	0.833808
TDEG-PRO	0.969106	RDKITS-2	0.925050	OWNBK-T?	0.972116
TDEG-?	0.968144	RDKITS-3	0.926226	PRJCT-T2	0.919312
UGRENG-?	0.817532	RDKITS-?	0.932001	PRJCT-T3	0.936186
UGRRED-?	0.802372	SOFTWR-2	0.856146	PRJCT-T4	0.885627
GRDENG-?	0.798481	SOFTWR-3	0.854938	PRJCT-T?	0.973177
GRDRED-?	0.864292	SOFTWR-4	0.933517	INTRP-T2	0.859158
READDEV2	0.893615	SOFTWR-?	0.942685	INTRP-T3	0.895705
READDEV3	0.910954	VARBKS-2	0.832131	INTRP-T4	0.824426
READDEV4	0.912961	VARBKS-3	0.879195	INTRP-T?	0.981055
READDEV5	0.904093	VARBKS-4	0.842370	EXPLA-T2	0.857884
READDEV5	0.982432	VARBKS-?	0.963420	EXPLA-T3	0.836140
PRPER-<1	0.926856	OTHMAT-2	0.912155	EXPLA-T4	0.890677
PRPER-12	0.949596	OTHMAT-3	0.926249	EXPLA-T?	0.989515
PRPER->2	0.960135	OTHMAT-4	0.905583	QUIZS-T2	0.918119
PRPER-?	0.986337	OTHMAT-?	0.939295	QUIZS-T3	0.948785
CLASIZ-2	0.885979	VOCAB-T2	0.805250	QUIZS-T4	0.827074
CLASIZ-3	0.899041	VOCAB-T3	0.861363	QUIZA-T?	0.982130
CLASIZ-4	0.903859	VOCAB-T4	0.847874	MOVIE-T2	0.917154
CLASIZ-5	0.942008	VOCAB-T?	0.988682	MOVIE-T3	0.966170
CLASIZ-?	0.830349	ALOUD-T2	0.842474	MOVIE-T4	0.946413
T_ABIL-N	0.789224	ALOUD-T3	0.873614	MOVIE-T?	0.989166
T_ABIL-?	0.965301	ALOUD-T4	0.815463	CMPTR-T2	0.859121
ABILRED2	0.895268	ALOUD-T?	0.989140	CMPTR-T3	0.875441
ABILRED3	0.899820	TLKRD-T2	0.849419	CMPTR-T4	0.945232
ABILRED4	0.917288	TLKRD-T3	0.901917	CMPTR-T?	0.988160
ABILRED?	0.970632	TLKRD-T4	0.872566	MCTEST-2	0.841895
INSTIME2	0.814598	TLKRD-T?	0.987528	MCTEST-3	0.869652
INSTIME3	0.922681	WRTRD-T2	0.841547	MCTEST-4	0.861380
INSTIME4	0.943764	WRTRD-T3	0.884369	MCTEST-?	0.958532
INSTIME?	0.928681	WRTRD-T4	0.870606	SATEST-2	0.833179
FLEX GRP	0.783369	WRTRD-T?	0.987307	SATEST-3	0.883440
2 GROUPS	0.907331	WB/WS-T2	0.894592	SATEST-4	0.814169
3 GROUPS	0.929888	WB/WS-T3	0.936419	SATEST-?	0.967341
4 GROUPS	0.961168	WB/WS-T4	0.920060	WRITST-2	0.796076

Table C-12 (continued)

Proportion of Variance of the Conditioning Variable Contrasts Accounted for by the Principal Components Used in the Conditioning Model for Reading Main Conditioning Variables, Age 13/Grade 8

Contrast	Proportion of Variance	Contrast	Proportion of Variance	Contrast	Proportion of Variance
WRITST-3	0.902387	PRTTST-3	0.890163	CLALIB-2	0.884740
WRITST-4	0.840450	PRTTST-4	0.892716	CLALIB-3	0.890876
WRITST-?	0.979514	PRTTST-?	0.957336	CLALIB-4	0.880509
PRJTST-2	0.917277	CLSLIB-2	0.862508	CLALIB-5	0.896791
PRJTST-3	0.928521	CLSLIB-3	0.918330	CLALIB-5	0.958224
PRJTST-4	0.863531	CLSLIB-4	0.882079	COMP-DIF	0.820260
PRJTST-?	0.964157	CLSLIB-5	0.850162	COMP-AVL	0.876854
PRTTST-2	0.906368	CLSLIB-?	0.979625	COMP-?	0.954114

Table C-13

Proportion of Variance of the Conditioning Variable Contrasts Accounted for by the Principal Components Used in the Conditioning Model for Reading Main Conditioning Variables, Age 17/Grade 12

Contrast	Proportion of Variance	Contrast	Proportion of Variance	Contrast	Proportion of Variance
FEMALE	0.930337	HW-YES	0.990005	SEMSCI-^	0.903643
BLACK	0.922538	HWLIN-0	0.980216	#SCI-LIN	0.869656
HISPANIC	0.972145	HWQUAD-0	0.980488	SEMHIS-^	0.892063
ASIAN	0.958822	HITEM=3	0.888299	#HIS-LIN	0.833813
PAC ISLD	0.973509	HITEM=4	0.858316	SEMFLG-^	0.886897
MEXICAN	0.960016	MOMHOM-N	0.965996	#FLG-LIN	0.741935
PUER RIC	0.985063	MOMHOM-?	0.861945	SEMVOC-^	0.888179
CUBN,OTH	0.973574	DADHOM-N	0.921454	#VOC-LIN	0.864147
HISP-?	0.986380	DADHOM-?	0.877798	SEMART-^	0.904218
NON MSA	0.813462	MISS-2<	0.985626	#ART-LIN	0.844351
MID CTY5	0.931929	USA 3-5	0.881193	G/R 22	0.914498
FR/BTWN5	0.928751	USA <3	0.842066	G/R 23	0.920653
SML TWN5	0.940272	USA-?	0.811369	G/R 24	0.965390
RURAL5	0.941967	STGRD1-2	0.973211	G/R 25	0.981063
URBAN FR	0.869390	STGRD3-5	0.947323	G/T 22	0.870092
MED CITY	0.785311	STGRD6>	0.933653	G/T 23	0.870084
SM PLACE	0.883808	SCHCHG-1	0.981191	G/T 24	0.678081
HS GRAD	0.962012	SCHCHG-2	0.988762	G/T 25	0.713722
POST HS	0.948676	SCHCHG-3	0.880834	G/P 22	0.911587
COL GRAD	0.941026	DIS@HOM2	0.918408	G/P 23	0.958770
PARED-?	0.962366	DIS@HOM3	0.927958	G/P 24	0.894044
S EAST	0.899034	DIS@HOM4	0.860591	G/P 25	0.871525
CENTRAL	0.908081	PGS>5	0.805089	G/S 22	0.942973
WEST	0.920048	PGS>10	0.813563	G/S 23	0.930904
PRIVATE	0.912442	COMP-WK	0.933625	R/T 22	0.895089
CATHOLIC	0.926765	COMP-MO	0.919724	R/T 23	0.877283
IEP-NO	0.986034	COMP-NEV	0.904167	R/T 24	0.861038
LEP-NO	0.839109	COMP-?	0.845383	R/T 25	0.877195
CHAP1-N	0.867008	ACADEMIC	0.737550	R/T 32	0.890113
INTEGRAT	0.911892	VOC/TECH	0.920726	R/T 33	0.903074
MINORITY	0.816237	OTHERPGM	0.953584	R/T 34	0.885271
NO LUNCH	0.750839	HS PGM-?	0.801013	R/T 35	0.899897
LUNCH-?	0.865843	VOC/BUSN	0.976361	R/T 42	0.946094
=MA/<MG	0.983594	2 YR COL	0.957424	R/T 43	0.947068
=MA/=MG	0.991008	4 YR COL	0.933081	R/T 44	0.948225
>MA/=MG	0.993311	MILITARY	0.982547	R/T 45	0.954899
HL-SOME	0.864430	OTHERACT	0.983282	R/T 52	0.973207
HL-ALWAY	0.763390	AFTRHS-?	0.780702	R/T 53	0.970369
HL-?	0.822373	SEMENG-^	0.884445	R/T 54	0.971467
TV-LIN	0.981512	#ENG-LIN	0.896836	R/T 55	0.959349
TV-QUAD	0.981527	SEMMAT-^	0.915221	R/P 22	0.901448
HW-NO	0.989664	#MAT-LIN	0.902670	R/P 23	0.905925

BEST COPY AVAILABLE

Table C-13 (continued)

Proportion of Variance of the Conditioning Variable Contrasts Accounted for by the Principal Components Used in the Conditioning Model for Reading Main Conditioning Variables, Age 17/Grade 12

Contrast	Proportion of Variance	Contrast	Proportion of Variance	Contrast	Proportion of Variance
R/P 24	0.902046	T/S 42	0.928621	RDMAGZ-4	0.936948
R/P 25	0.830217	T/S 43	0.938557	RDMAGZ-?	0.695227
R/P 32	0.892989	T/S 52	0.924186	VOCAB-2	0.917985
R/P 33	0.903622	T/S 53	0.944887	VOCAB-3	0.932339
R/P 34	0.912061	P/S 22	0.977563	VOCAB-4	0.921511
R/P 35	0.832229	P/S 23	0.958375	VOCAB-?	0.885851
R/P 42	0.964959	P/S 32	0.961699	TALKRD-2	0.952985
R/P 43	0.969000	P/S 33	0.946369	TALKRD-3	0.951227
R/P 44	0.957569	P/S 42	0.958887	TALKRD-4	0.940174
R/P 45	0.903377	P/S 43	0.936111	TALKRD-?	0.806407
R/P 52	0.978026	P/S 52	0.983309	WBK/WS-2	0.928621
R/P 53	0.979963	P/S 53	0.967659	WBK/WS-3	0.936379
R/P 54	0.977772	NBOOK-12	0.877700	WBK/WS-4	0.917652
R/P 55	0.964706	NBOOK-34	0.907660	WBK/WS-?	0.645769
R/S 22	0.976297	NBOOK-5+	0.936110	WRTRED-2	0.923778
R/S 23	0.969536	NBOOK-?	0.906235	WRTRED-3	0.912429
R/S 32	0.974483	GOOD RDR	0.916779	WRTRED-4	0.937416
R/S 33	0.960849	AVG RDR	0.892856	WRTRED-?	0.738685
R/S 42	0.973264	POOR RDR	0.954115	RDPRJ-S2	0.963764
R/S 43	0.975899	RDR-MISS	0.760518	RDPRJ-S3	0.963997
R/S 52	0.982992	RD4FUN-2	0.942544	RDPRJ-S4	0.965552
R/S 53	0.975739	RD4FUN-3	0.927613	RDPRJ-S?	0.824380
T/P 22	0.813005	RD4FUN-4	0.919376	ALOUD-S2	0.935358
T/P 23	0.783502	RD4FUN-?	0.900374	ALOUD-S3	0.933708
T/P 24	0.744562	FAMRED-2	0.933305	ALOUD-S4	0.931878
T/P 25	0.906757	FAMRED-3	0.940432	ALOUD-S?	0.669632
T/P 32	0.745948	FAMRED-4	0.942016	SILNT-S2	0.864255
T/P 33	0.714993	FAMRED-?	0.889525	SILNT-S3	0.867978
T/P 34	0.708119	NOVEL-2	0.957711	SILNT-S4	0.917394
T/P 35	0.889822	NOVEL-3	0.945609	SILNT-S?	0.706263
T/P 42	0.808019	NOVEL-4	0.946712	OWNBK-S2	0.973544
T/P 43	0.805668	NOVEL-?	0.824617	OWNBK-S3	0.964376
T/P 44	0.808472	USELIB-2	0.954231	OWNBK-S4	0.955879
T/P 45	0.926915	USELIB-3	0.969757	OWNBK-S?	0.768581
T/P 52	0.795808	USELIB-4	0.948142	EXPLA-2	0.839134
T/P 53	0.804564	USELIB-?	0.908699	EXPLA-3	0.930042
T/P 54	0.807971	RDNEWS-2	0.892293	EXPLA-4	0.790598
T/P 55	0.929596	RDNEWS-3	0.891442	EXPLA-?	0.847008
T/S 22	0.915448	RDNEWS-4	0.886411	DISCU-2	0.856066
T/S 23	0.910823	RDNEWS-?	0.674353	DISCU-3	0.931578
T/S 32	0.917409	RDMAGZ-2	0.875344	DISCU-4	0.825602
T/S 33	0.894265	RDMAGZ-3	0.871852	DISCU-?	0.869719

Table C-13 (continued)

Proportion of Variance of the Conditioning Variable Contrasts Accounted for by the Principal Components Used in the Conditioning Model for Reading Main Conditioning Variables, Age 17/Grade 12

Contrast	Proportion of Variance	Contrast	Proportion of Variance	Contrast	Proportion of Variance
AC2LIB-N	0.977267	TESTEFF?	0.885029	SCH%RURL	0.813490
AC2LIB-?	0.797139	TESTIMP2	0.957833	RUR%-LIN	0.815326
LIB4RS-2	0.890521	TESTIMP3	0.950048	SCH%CITY	0.799269
LIB4RS-3	0.955167	TESTIMP4	0.946173	CITY%-LN	0.821675
LIB4RS-4	0.948713	TESTIMP?	0.877171	PAR%PROF	0.960240
LIB4RS-5	0.894135	LNGANSW2	0.851339	PROF%-LN	0.823312
LIB4RS-?	0.877611	LNGANSW3	0.854599	PAR%BCOL	0.952876
LIB2BW-2	0.880873	LNGANSW4	0.910153	BCOL%-LN	0.837539
LIB2BW-3	0.954523	LNGANSW?	0.857366	PAR%FARM	0.856795
LIB2BW-4	0.953798	SNRM-LIN	0.749435	FARM%-LN	0.757629
LIB2BW-5	0.901868	ENGABL-N	0.905736	PAR%IRRE	0.918046
LIB2BW-?	0.887018	ENGABL-?	0.911912	IRRE%-LN	0.761209
LIB4U-2	0.934349	%SUBLUN2	0.920040	PAR%WELF	0.892775
LIB4U-3	0.953001	%SUBLUN3	0.894991	WELF%-LN	0.801343
LIB4U-4	0.975073	%SUBLUN4	0.907391	LSTAF-NO	0.917464
LIB4U-5	0.926194	%SUBLUN5	0.863410	LSTAF-PT	0.965950
LIB4U-?	0.889163	%SUBLUN?	0.854722	LSTAF-FT	0.963667
LIB4NJ-2	0.894218	%REMRED2	0.884482	LSTAF-?	0.948829
LIB4NJ-3	0.920563	%REMRED3	0.853956	#AP ENG2	0.930419
LIB4NJ-4	0.981079	%REMRED4	0.860194	#AP ENG3	0.957321
LIB4NJ-5	0.875505	%REMRED5	0.871704		
LIB4NJ-?	0.902049	%REMRED?	0.913644		
LIB4QI-2	0.950221	%ENREOY2	0.889136		
LIB4QI-3	0.941690	%ENREOY3	0.916856		
LIB4QI-4	0.945242	%ENREOY4	0.881438		
LIB4QI-5	0.896761	%ENREOY?	0.901534		
LIB4QI-?	0.912323	PARAID-O	0.959185		
PFOLIO-N	0.976028	PARAID-N	0.963331		
PFOLIO-?	0.823379	PARAID-?	0.955277		
AP ENG-N	0.934370	PARHW-N?	0.911579		
AP ENG-?	0.806850	PARHW-N?	0.895906		
#QUEST+2	0.922442	S/PHW-N?	0.969916		
#QUEST+3	0.903455	S/PHW-N?	0.971979		
#QUEST+4	0.926788	VOLPG-N?	0.909792		
#QUEST+?	0.813742	VOLPG-N?	0.823152		
TESTDIF2	0.979522	SCHCH1-N	0.807238		
TESTDIF3	0.968487	SCHCH1-?	0.920312		
TESTDIF4	0.967493	%CHAP1-2	0.898369		
TESTDIF?	0.884756	%CHAP1-3	0.868743		
TESTEFF2	0.981483	%CHAP1-4	0.895034		
TESTEFF3	0.965359	%CHAP1-5	0.909736		
TESTEFF4	0.967636	%CHAP1-?	0.751262		

Table C-14

Proportion of Variance of the Conditioning Variable Contrasts Accounted for by the Principal Components Used in the Conditioning Model for U.S. History Conditioning Variables, Age 9/Grade 4

Contrast	Proportion of Variance	Contrast	Proportion of Variance	Contrast	Proportion of Variance
FEMALE	0.927800	HW-NO	0.988991	G/S 23	0.896725
BLACK	0.930056	HW-YES	0.988979	R/T 25	0.908706
HISPANIC	0.969852	HWLIN-0	0.987062	R/T 31	0.917599
ASIAN	0.966889	HWQUAD-0	0.985771	R/T 32	0.905535
PAC ISLD	0.985915	HITEM=3	0.846000	R/T 33	0.926280
MEXICAN	0.957572	HITEM=4	0.808902	R/T 34	0.900237
PUER RIC	0.965459	MOMHOM-N	0.900177	R/T 35	0.900384
CUBN,OTH	0.975004	MOMHOM-?	0.752397	R/T 41	0.892816
HISP-?	0.936806	DADHOM-N	0.873500	R/T 42	0.895141
NON MSA	0.873871	DADHOM-?	0.799388	R/T 43	0.958203
MID CTY5	0.937300	MISS-2<	0.971755	R/T 44	0.959812
FR/BTWN5	0.938256	USA 3-5	0.959182	R/T 45	0.965093
SML TWN5	0.943260	USA <3	0.965639	R/T 51	0.961979
RURAL5	0.952165	USA-?	0.768408	R/T 52	0.971323
URBAN FR	0.873996	STGRD1-2	0.882916	R/T 53	0.971286
MED CITY	0.843320	STGRD3>	0.888921	R/T 54	0.962698
SM PLACE	0.905372	PRESCH-N	0.945926	R/T 55	0.968397
HS GRAD	0.949230	SCHCHG-1	0.880755	R/P 25	0.897790
POST HS	0.930924	SCHCHG-2	0.889382	R/P 31	0.882996
COL GRAD	0.930883	SCHCHG-3	0.833445	R/P 32	0.912423
PARED-?	0.936890	DIS@HOM2	0.884210	R/P 33	0.908438
S EAST	0.857814	DIS@HOM3	0.914989	R/P 34	0.896821
CENTRAL	0.839283	DIS@HOM4	0.815675	R/P 35	0.906208
WEST	0.875506	PGS>5	0.811742	R/P 41	0.914975
PRIVATE	0.919603	PGS>10	0.822445	R/P 42	0.907468
CATHOLIC	0.929549	COMP-WK	0.940826	R/P 43	0.981916
IEP-NO	0.963808	COMP-MO	0.955706	R/P 44	0.977879
LEP-NO	0.868747	COMP-NEV	0.929516	R/P 45	0.978809
CHAP1-N	0.880106	COMP-?	0.747787	R/P 51	0.977035
INTEGRAT	0.864194	G/R 22	0.915770	R/P 52	0.983653
MINORITY	0.843477	G/R 23	0.915533	R/P 53	0.976035
NO LUNCH	0.824683	G/R 24	0.976749	R/P 54	0.985621
LUNCH-?	0.829730	G/R 25	0.984734	R/P 55	0.985400
=MA/<MG	0.869411	G/T 22	0.867011	R/S 32	0.963158
=MA/=MG	0.956015	G/T 23	0.815245	R/S 33	0.962394
=MA/>MG	0.991370	G/T 24	0.721833	R/S 41	0.970774
>MA/=MG	0.962116	G/T 25	0.741241	R/S 42	0.950577
HL-SOME	0.910390	G/P 22	0.931959	R/S 43	0.977609
HL-ALWAY	0.926362	G/P 23	0.914960	R/S 51	0.985148
HL-?	0.805713	G/P 24	0.858216	R/S 52	0.985508
TVLIN-0	0.984321	G/P 25	0.849901	R/S 53	0.983646
TV-QUAD	0.985328	G/S 22	0.918261	T/P 25	0.813652

Table C-14 (continued)

Proportion of Variance of the Conditioning Variable Contrasts Accounted for by the Principal Components Used in the Conditioning Model for U.S. History Conditioning Variables, Age 9/Grade 4

Contrast	Proportion of Variance	Contrast	Proportion of Variance	Contrast	Proportion of Variance
T/P 31	0.852969	RDTEXT-4	0.933838	MAPSGL-?	0.728297
T/P 32	0.801632	RDTEXT-5	0.872835	VIDEO--2	0.916067
T/P 33	0.800599	RDTEXT-?	0.758539	VIDEO--3	0.909869
T/P 34	0.778322	EXTMAT-2	0.946399	VIDEO--4	0.939712
T/P 35	0.800370	EXTMAT-3	0.960323	VIDEO--5	0.959819
T/P 41	0.782957	EXTMAT-4	0.962314	VIDEO--?	0.749021
T/P 42	0.782865	EXTMAT-5	0.948870	COMPUT-2	0.914150
T/P 43	0.836368	EXTMAT-?	0.761534	COMPUT-3	0.943205
T/P 44	0.886967	HISPEO-2	0.960220	COMPUT-4	0.946208
T/P 45	0.782344	HISPEO-3	0.959880	COMPUT-5	0.912525
T/P 51	0.782332	HISPEO-4	0.963843	COMPUT-?	0.722373
T/P 52	0.803248	HISPEO-5	0.947868	QUIZ---2	0.924610
T/P 53	0.860134	HISPEO-?	0.768930	QUIZ---3	0.930387
T/P 54	0.798602	DISMAT-2	0.926934	QUIZ---4	0.958127
T/P 55	0.779061	DISMAT-3	0.937057	QUIZ---5	0.973908
T/S 32	0.915975	DISMAT-4	0.938547	QUIZ---?	0.799209
T/S 33	0.905552	DISMAT-5	0.920393	TRIPS--2	0.976069
T/S 41	0.912208	DISMAT-?	0.856924	TRIPS--3	0.965199
T/S 42	0.904228	WSHANS-2	0.887372	TRIPS--4	0.961507
T/S 43	0.925936	WSHANS-3	0.910035	TRIPS--5	0.963209
T/S 51	0.940760	WSHANS-4	0.938494	TRIPS--?	0.841878
T/S 52	0.947982	WSHANS-5	0.923181	LIBRA--2	0.957190
T/S 53	0.936485	WSHANS-?	0.761061	LIBRA--3	0.969172
P/S 32	0.963237	WREPRT-2	0.964563	LIBRA--4	0.973180
P/S 33	0.951326	WREPRT-3	0.953767	LIBRA--5	0.965984
P/S 41	0.934424	WREPRT-4	0.949999	LIBRA--?	0.840983
P/S 42	0.923151	WREPRT-5	0.970950	#QUEST+2	0.879514
P/S 43	0.947786	WREPRT-?	0.810875	#QUEST+3	0.879692
P/S 51	0.934213	GRPPRJ-2	0.941747	#QUEST+4	0.914645
P/S 52	0.949463	GRPPRJ-3	0.942213	#QUEST+?	0.821566
P/S 53	0.934262	GRPPRJ-4	0.949764	TESTDIF2	0.868330
OFT_3-4W	0.907175	GRPPRJ-5	0.953838	TESTDIF3	0.874767
OFT_1-2W	0.893262	GRPPRJ-?	0.723013	TESTDIF4	0.890754
OFT_<1W	0.933029	GIVREP-2	0.962422	TESTDIF?	0.896934
OFT_NEV	0.913988	GIVREP-3	0.951554	TESTEFF2	0.871371
OFT_MIS	0.699894	GIVREP-4	0.949806	TESTEFF3	0.857825
NO	0.953022	GIVREP-5	0.944380	TESTEFF4	0.911249
MISSING	0.823667	GIVREP-?	0.715041	TESTEFF?	0.881803
PORT--NO	0.971009	MAPSGL-2	0.918069	TESTIMP2	0.913411
PORT-MIS	0.805688	MAPSGL-3	0.923519	TESTIMP3	0.921093
RDTEXT-2	0.926422	MAPSGL-4	0.938056	TESTIMP4	0.937820
RDTEXT-3	0.942144	MAPSGL-5	0.956025	TESTIMP?	0.936977

Table C-14 (continued)

Proportion of Variance of the Conditioning Variable Contrasts Accounted for by the Principal Components Used in the Conditioning Model for U.S. History Conditioning Variables, Age 9/Grade 4

Contrast	Proportion of Variance	Contrast	Proportion of Variance	Contrast	Proportion of Variance
LNGANSW2	0.871334	SCH%RURL	0.818413	CERTR-N	0.901278
LNGANSW3	0.893232	RUR%-LIN	0.847618	CERTR-NS	0.856387
LNGANSW4	0.865848	SCH%CITY	0.836536	CERTR-?	0.918026
LNGANSW?	0.939555	CITY%-LN	0.882204	CERTL-N	0.887641
SNRM-LIN	0.812804	PAR%PROF	0.929297	CERTL-NS	0.860015
DEPTLIZD	0.833603	PROF%-LN	0.858850	CERTL-?	0.897569
REGROUPD	0.875446	PAR%BCOL	0.942648	RESOURC2	0.911052
SCH4ORG?	0.871384	BCOL%-LN	0.829831	RESOURC3	0.911850
CLAABL-N	0.809878	PAR%FARM	0.868852	RESOURC4	0.925210
CLAABL-?	0.893976	FARM%-LN	0.799955	RESOURC?	0.988260
RD PRI-N	0.853554	PAR%IRRE	0.910439	UGR ED-Y	0.813747
RD PRI-?	0.921513	IRRE%-LN	0.819684	GRA ED-Y	0.870635
%SUBLUN2	0.882120	PAR%WELF	0.923858	NOGRAD-Y	0.884983
%SUBLUN3	0.910293	WELF%-LN	0.867093	T4REDYR2	0.904696
%SUBLUN4	0.898833	LSTAF-NO	0.909249	T4REDYR3	0.909074
%SUBLUN5	0.874440	LSTAF-PT	0.951276	T4REDYR4	0.934877
%SUBLUN?	0.914488	LSTAF-FT	0.952934	T4REDYR5	0.951531
%REMRED2	0.876700	CMPCL-N?	0.833550	T4REDYR?	0.966953
%REMRED3	0.886643	CMPLB-N?	0.819568	TWKSP-<6	0.856149
%REMRED4	0.861720	CMPBR-N?	0.847751	TWKSP-15	0.865540
%REMRED5	0.837922	HS PRI-N	0.851418	TWKSP-35	0.871313
%REMRED?	0.898850	HS PRI-?	0.908130	TWKSP>35	0.866971
%ENREOY2	0.877194	GE PRI-N	0.832665	TWKSP-?	0.978446
%ENREOY3	0.898330	GE PRI-?	0.918211	CERT-TMP	0.894354
%ENREOY4	0.896287	TMCH-PAR	0.826340	CERT-REG	0.934305
%ENREOY?	0.918682	TMCH-COM	0.981851	CERT-HGH	0.947999
PARAID-O	0.864070	T_FEMALE	0.835617	CERT-?	0.965527
PARAID-N	0.904653	T_SEX-?	0.987456	TDEG-ASC	0.948922
PARAID-?	0.935846	T_BLACK	0.839306	TDEG-BAC	0.933097
PARHW-OC	0.829637	T_HISP	0.830717	TDEG-MAS	0.923061
PARHW-N?	0.929606	T_ASIAN	0.897837	TDEG-EDS	0.908565
S/PHW-OC	0.850357	T_PAC IS	0.829761	TDEG-?	0.978697
S/PHW-N?	0.821523	T_AM IND	0.847315	UGRENG-?	0.828709
VOLPG-OC	0.890190	T_RACE-?	0.962646	UGRRED-?	0.857071
VOLPG-N?	0.798082	T_YREXP2	0.890907	GRDENG-?	0.829700
SCHCH1-N	0.826142	T_YREXP3	0.887787	GRDRED-?	0.856163
SCHCH1-?	0.956376	T_YREXP4	0.924208	READDEV2	0.898752
%CHAP1-2	0.893359	T_YREXP5	0.934459	READDEV3	0.917611
%CHAP1-3	0.895608	T_YREXP?	0.989534	READDEV4	0.896902
%CHAP1-4	0.865578	CERTG-N	0.844182	READDEV5	0.895116
%CHAP1-5	0.842081	CERTG-NS	0.973102	READDEV5	0.984999
%CHAP1-?	0.818269	CERTG-?	0.965353	PRPER-<1	0.887806

Table C-14 (continued)

Proportion of Variance of the Conditioning Variable Contrasts Accounted for by the Principal Components Used in the Conditioning Model for U.S. History Conditioning Variables, Age 9/Grade 4

Contrast	Proportion of Variance	Contrast	Proportion of Variance	Contrast	Proportion of Variance
PRPER-12	0.923755	CLASAB-N	0.901618	EMPWRL-L	0.909141
PRPER->2	0.925236	CLASAB-?	0.979596	EMPWRL-?	0.965151
PRPER-?	0.970024	CLASIZ-A	0.943274	HISHWK-H	0.845286
CSPECS-N	0.805358	CLASIZ-L	0.913339	HISHWK-1	0.865054
CSPECS-?	0.952106	CLASIZ-M	0.932710	HISHWK-2	0.875853
CERTH-N	0.931078	CLASIZ-?	0.982653	HISHW-2+	0.868304
CERTH-NS	0.929877	FOC-US-H	0.826821	HISHWK-?	0.977020
CERTH-?	0.925882	FOC-SS	0.854878	RDTXT--W	0.844622
CERTG-N	0.939668	FOC-SS+	0.852940	RDTXT--M	0.860015
CERTG-NS	0.913651	FOC-GEOG	0.882473	RDTXT--N	0.845035
CERTG-?	0.950411	FOC-OTH	0.869934	RDTXT--?	0.991605
CERTO-N	0.822563	FOC-MISS	0.891019	RDOTH--W	0.889869
CERTO-NS	0.860373	TIME-45M	0.894457	RDOTH--M	0.922250
CERTO-?	0.866017	TIME-60M	0.877310	RDOTH--N	0.880762
UGR GE-Y	0.853221	TIME-90+	0.854601	RDOTH--?	0.990596
UGR HI-Y	0.805243	TIME-MIS	0.967876	USEPRM-W	0.853643
UGR SS-Y	0.857629	WK1815-2	0.848289	USEPRM-M	0.912840
UGR OT-Y	0.833470	WK1815-3	0.847771	USEPRM-N	0.920930
GRD GE-Y	0.824599	WK1815-4	0.862929	USEPRM-?	0.985654
GRD HI-Y	0.862293	WK1815-?	0.984774	TESTS--W	0.901165
GRD SS-Y	0.813042	WK1865-2	0.852101	TESTS--M	0.950496
GRA OT-Y	0.858447	WK1865-3	0.877533	TESTS--N	0.881136
PORT---N	0.842903	WK1865-4	0.899591	TESTS--?	0.978932
PORT---?	0.979635	WK1865-?	0.980261	DISCUS-W	0.801089
PORTR--N	0.808743	WK1945-2	0.841061	DISCUS-M	0.872094
PORTR--?	0.977810	WK1945-3	0.858465	DISCUS-N	0.836766
T4HISYR2	0.863492	WK1945-4	0.901002	DISCUS-?	0.989598
T4HISYR3	0.863774	WK1945-?	0.974991	WRTSHT-W	0.917941
T4HISYR4	0.892676	WKCURR-2	0.854330	WRTSHT-M	0.931531
T4HISYR5	0.927253	WKCURR-3	0.882817	WRTSHT-N	0.877246
T4HISYR?	0.974831	WKCURR-4	0.890870	WRTSHT-?	0.987066
TWKSG-<6	0.836231	WKCURR-?	0.973749	WRTRPT-W	0.859663
TWKSG-15	0.835854	EMPDEM-M	0.869302	WRTRPT-M	0.940061
TWKSG-35	0.879822	EMPDEM-L	0.894685	WRTRPT-N	0.895859
TWKSG>35	0.866009	EMPDEM-?	0.972038	GRPPRJ-W	0.842584
TWKSG-?	0.984379	EMPCUL-M	0.805363	GRPPRJ-M	0.942156
CSPECG-N	0.944487	EMPCUL-L	0.850602	GRPPRJ-N	0.875706
CSPECG-?	0.927016	EMPCUL-?	0.972559	GRPPRJ-?	0.989873
CSPECS-N	0.926063	EMPTEC-M	0.859084	ORALRP-W	0.829106
CSPECS-?	0.927150	EMPTEC-L	0.870461	ORALRP-M	0.927851
CSPECH-N	0.948380	EMPTEC-?	0.970791	ORALRP-N	0.899177
CSPECH-?	0.919897	EMPWRL-M	0.904526	ORALRP-?	0.979673

BEST COPY AVAILABLE

Table C-14 (continued)

Proportion of Variance of the Conditioning Variable Contrasts Accounted for by the Principal Components Used in the Conditioning Model for U.S. History Conditioning Variables, Age 9/Grade 4

Contrast	Proportion of Variance	Contrast	Proportion of Variance	Contrast	Proportion of Variance
USEMAP-W	0.829077	LIBOOK-M	0.911774	CL436+	0.883019
USEMAP-M	0.881937	LIBOOK-N	0.890777	CL4-MISS	0.975655
USEMAP-N	0.847409	LIBOK-NO	0.880059	PRPHIS-A	0.870334
USEMAP-?	0.990446	LIBOOK-?	0.979709	PRPHIS-S	0.888889
VIDEOS-W	0.893556	COMPAV-D	0.865001	PRPHIS-U	0.856248
VIDEOS-M	0.952993	COMPAV-C	0.876480	PRPHIS-?	0.951104
VIDEOS-N	0.918801	COMPAV-?	0.978358	PRPSS--A	0.850286
VIDEOS-?	0.988732	MCTEST-M	0.905617	PRPSS--S	0.860454
COMPUT-W	0.898346	MCTEST-Y	0.916636	PRPSS--U	0.866184
COMPUT-M	0.918513	MCTEST-N	0.870939	PRPSS--?	0.956817
COMPUT-N	0.933881	MCTEST-?	0.978806	UNICOU-Y	0.896666
COMPUT-?	0.990541	WRTRSP-M	0.834928	WRKSHP-Y	0.847993
TRIPS--W	0.840349	WRTRSP-Y	0.900621	PROACT-Y	0.853639
TRIPS--M	0.940439	WRTRSP-N	0.843337	NOED5--Y	0.897557
TRIPS--N	0.958485	WRTRSP-?	0.973616	UNICOS-Y	0.894020
TRIPS--?	0.990605	GRPPRJ-M	0.897949	WRKSHS-Y	0.874888
WKLIBR-W	0.874926	GRPPRJ-Y	0.917810	PROACS-Y	0.871154
WKLIBR-M	0.921607	GRPPRJ-N	0.864055	NOED5S-Y	0.886248
WKLIBR-N	0.912096	GRPPRJ-?	0.981933	CRLE-DST	0.862428
WKLIBR-?	0.988630	TPORT--M	0.887862	CRLE-SCH	0.870125
GOLIB--W	0.929661	TPORT--Y	0.898868	CRLE-IND	0.829122
GOLIB--M	0.904576	TPORT--N	0.918179	CRLE-STA	0.853035
GOLIB--N	0.872598	TPORT--?	0.972630	CRLE-?	0.835333
GOLIB-NO	0.870244	CL421-25	0.869995	LKHIS-OT	0.816498
GOLIBR-?	0.991509	CL426-30	0.893021	LKHIS-NO	0.836580
LIBOOK-W	0.892867	CL431-35	0.852210	LKHIS-?	0.957301

Table C-15

Proportion of Variance of the Conditioning Variable Contrasts Accounted for by the Principal Components Used in the Conditioning Model for U.S. History Conditioning Variables, Age 13/Grade 8

Contrast	Proportion of Variance	Contrast	Proportion of Variance	Contrast	Proportion of Variance
FEMALE	0.923514	HW-NO	0.979683	G/S 23	0.939877
BLACK	0.921919	HW-YES	0.979748	R/T 25	0.902244
HISPANIC	0.969709	HWLIN-0	0.978912	R/T 31	0.898015
ASIAN	0.964967	HWQUAD-0	0.977912	R/T 32	0.908063
PAC ISLD	0.969614	HITEM=3	0.880834	R/T 33	0.898278
MEXICAN	0.958613	HITEM=4	0.842480	R/T 34	0.898513
PUER RIC	0.971854	MOMHOM-N	0.976502	R/T 35	0.911619
CUBN,OTH	0.976926	MOMHOM-?	0.875519	R/T 41	0.884237
HISP-?	0.978566	DADHOM-N	0.943574	R/T 42	0.912442
NON MSA	0.853940	DADHOM-?	0.907758	R/T 43	0.955667
MID CTY5	0.930391	MISS-2<	0.977180	R/T 44	0.953542
FR/BTWN5	0.934748	USA 3-5	0.933777	R/T 45	0.958677
SML TWN5	0.933392	USA <3	0.962308	R/T 51	0.965181
RURAL5	0.944860	USA-?	0.809215	R/T 52	0.973608
URBAN FR	0.867386	STGRD1-2	0.960943	R/T 53	0.972928
MED CITY	0.842480	STGRD3-5	0.938021	R/T 54	0.966060
SM PLACE	0.899351	STGRD6>	0.912809	R/T 55	0.971296
HS GRAD	0.956671	SCHCHG-1	0.926707	R/P 25	0.906567
POST HS	0.940206	SCHCHG-2	0.940688	R/P 31	0.900115
COL GRAD	0.936266	SCHCHG-3	0.760378	R/P 32	0.915310
PARED-?	0.957398	DIS@HOM2	0.910694	R/P 33	0.863531
S EAST	0.880222	DIS@HOM3	0.936165	R/P 34	0.907587
CENTRAL	0.871249	DIS@HOM4	0.819467	R/P 35	0.907022
WEST	0.864728	PGS>5	0.796211	R/P 41	0.924656
PRIVATE	0.916861	PGS>10	0.800371	R/P 42	0.872985
CATHOLIC	0.914526	COMP-WK	0.955021	R/P 43	0.983405
IEP-NO	0.973306	COMP-MO	0.946899	R/P 44	0.975135
LEP-NO	0.865174	COMP-NEV	0.936958	R/P 45	0.976874
CHAP1-N	0.819312	COMP-?	0.825399	R/P 51	0.960895
INTEGRAT	0.876173	G/R 22	0.911259	R/P 52	0.985783
MINORITY	0.836754	G/R 23	0.909664	R/P 53	0.979271
NO LUNCH	0.771943	G/R 24	0.972531	R/P 54	0.985109
LUNCH-?	0.867168	G/R 25	0.981218	R/P 55	0.980323
=MA/<MG	0.809426	G/T 22	0.835087	R/S 32	0.964472
=MA/=MG	0.943456	G/T 23	0.845150	R/S 33	0.971509
=MA/>MG	0.969164	G/T 24	0.702449	R/S 41	0.973524
>MA/=MG	0.946772	G/T 25	0.704584	R/S 42	0.945009
HL-SOME	0.862610	G/P 22	0.911267	R/S 43	0.977645
HL-ALWAY	0.796598	G/P 23	0.919182	R/S 51	0.977461
HL-?	0.825314	G/P 24	0.860366	R/S 52	0.982282
TVLIN-0	0.985924	G/P 25	0.744996	R/S 53	0.975450
TV-QUAD	0.985878	G/S 22	0.950574	T/P 25	0.789406

Table C-15 (continued)

Proportion of Variance of the Conditioning Variable Contrasts Accounted for by the Principal Components Used in the Conditioning Model for U.S. History Conditioning Variables, Age 13/Grade 8

Contrast	Proportion of Variance	Contrast	Proportion of Variance	Contrast	Proportion of Variance
T/P 31	0.804925	EXTMAT-?	0.785572	COMPUT-3	0.967371
T/P 32	0.745087	HISPEO-2	0.974630	COMPUT-4	0.966798
T/P 33	0.842595	HISPEO-3	0.975230	COMPUT-5	0.930474
T/P 34	0.743211	HISPEO-4	0.973734	COMPUT-?	0.793187
T/P 35	0.763639	HISPEO-5	0.964129	QUIZ---2	0.901608
T/P 41	0.715176	HISPEO-?	0.744213	QUIZ---3	0.913309
T/P 42	0.826741	DISMAT-2	0.928974	QUIZ---4	0.962641
T/P 43	0.821997	DISMAT-3	0.934010	QUIZ---5	0.960358
T/P 44	0.840654	DISMAT-4	0.926141	QUIZ---?	0.848800
T/P 45	0.807771	DISMAT-5	0.894446	TRIPS--2	0.981266
T/P 51	0.888360	DISMAT-?	0.711768	TRIPS--3	0.979352
T/P 52	0.791993	WSHANS-2	0.887256	TRIPS--4	0.982380
T/P 53	0.818104	WSHANS-3	0.900761	TRIPS--5	0.970724
T/P 54	0.804057	WSHANS-4	0.943143	TRIPS--?	0.880344
T/P 55	0.877869	WSHANS-5	0.911803	LIBRA--2	0.974397
T/S 32	0.908546	WSHANS-?	0.763936	LIBRA--3	0.976427
T/S 33	0.916424	WREPRT-2	0.927888	LIBRA--4	0.978257
T/S 41	0.928931	WREPRT-3	0.936392	LIBRA--5	0.968603
T/S 42	0.906522	WREPRT-4	0.967124	LIBRA--?	0.875161
T/S 43	0.925938	WREPRT-5	0.857133	HISGR6-M	0.820052
T/S 51	0.930079	WREPRT-?	0.725478	HISGR6-D	0.827642
T/S 52	0.931048	GRPPRJ-2	0.964809	HISGR6-?	0.875441
T/S 53	0.931986	GRPPRJ-3	0.965603	HISGR7-M	0.796073
P/S 32	0.966820	GRPPRJ-4	0.970042	HISGR7-D	0.826981
P/S 33	0.956984	GRPPRJ-5	0.946287	HISGR7-?	0.893166
P/S 41	0.951184	GRPPRJ-?	0.738873	HISNOW-M	0.878015
P/S 42	0.936149	GIVREP-2	0.939018	HISNOW-?	0.805307
P/S 43	0.952941	GIVREP-3	0.927404	HISHWK-2	0.937659
P/S 51	0.941708	GIVREP-4	0.968257	HISHWK-3	0.941182
P/S 52	0.970814	GIVREP-5	0.857149	HISHWK-4	0.955356
P/S 53	0.962484	GIVREP-?	0.756798	HISHWK-5	0.956725
PORT--NO	0.971586	MAPSGL-2	0.930651	HISHWK-?	0.884261
PORT-MIS	0.813299	MAPSGL-3	0.933006	PR1815-2	0.845974
RDTEXT-2	0.911178	MAPSGL-4	0.949068	PR1815-3	0.829515
RDTEXT-3	0.929151	MAPSGL-5	0.958016	PR1815-?	0.899582
RDTEXT-4	0.920060	MAPSGL-?	0.748306	181565-2	0.847612
RDTEXT-5	0.915067	VIDEO--2	0.951267	181565-3	0.816889
RDTEXT-?	0.825536	VIDEO--3	0.948918	181565-?	0.871108
EXTMAT-2	0.953693	VIDEO--4	0.962646	186545-2	0.866877
EXTMAT-3	0.960614	VIDEO--5	0.957482	186545-3	0.835221
EXTMAT-4	0.966745	VIDEO--?	0.826283	186545-?	0.880176
EXTMAT-5	0.956189	COMPUT-2	0.965775	PS1945-2	0.878627

Table C-15 (continued)

Proportion of Variance of the Conditioning Variable Contrasts Accounted for by the Principal Components Used in the Conditioning Model for U.S. History Conditioning Variables, Age 13/Grade 8

Contrast	Proportion of Variance	Contrast	Proportion of Variance	Contrast	Proportion of Variance
PS1945-3	0.871060	%ENREOY4	0.889196	HS PRI-?	0.910051
PW1945-?	0.889035	%ENREOY?	0.914558	GE PRI-N	0.861117
#QUEST+2	0.935993	PARAID-O	0.883541	GE PRI-?	0.906045
#QUEST+3	0.936335	PARAID-N	0.911045	HTSUB#=1	0.908490
#QUEST+4	0.951174	PARAID-?	0.980446	NO HIST8	0.889668
#QUEST+?	0.928434	PARHW-OC	0.846190	HTSUB#-?	0.968711
TESTDIF2	0.904819	PARHW-N?	0.845431	GTSUB#=1	0.930823
TESTDIF3	0.905242	S/PHW-OC	0.904736	NO GEOG8	0.904838
TESTDIF4	0.925881	S/PHW-N?	0.926902	GTSUB#-?	0.902546
TESTDIF?	0.921164	VOLPG-OC	0.845629	HISABL-N	0.900172
TESTEFF2	0.927890	VOLPG-N?	0.816393	HISABL-?	0.934402
TESTEFF3	0.915585	SCHCH1-N	0.831954	GEOABL-N	0.894779
TESTEFF4	0.934764	SCHCH1-?	0.958897	GEOABL-?	0.853843
TESTEFF?	0.930284	%CHAP1-2	0.868705	TMCH-PAR	0.845025
TESTIMP2	0.918385	%CHAP1-3	0.856255	TMCH-COM	0.980051
TESTIMP3	0.910434	%CHAP1-4	0.864818	T_FEMALE	0.780746
TESTIMP4	0.908598	%CHAP1-5	0.855786	T_SEX-?	0.981832
TESTIMP?	0.954437	%CHAP1-?	0.786745	T_BLACK	0.862542
LNGANSW2	0.874137	SCH%RURL	0.814915	T_HISP	0.904364
LNGANSW3	0.881314	RUR%-LIN	0.877224	T_ASIAN	0.903998
LNGANSW4	0.882446	SCH%CITY	0.821175	T_PAC IS	0.884727
LNGANSW?	0.954706	CITY%-LN	0.862439	T_RACE-?	0.973518
SNRM-LIN	0.818309	PAR%PROF	0.956900	T_YREXP2	0.870469
SEMIDEPT	0.909977	PROF%-LN	0.852809	T_YREXP3	0.885411
DEPTLIZD	0.913766	PAR%BCOL	0.958714	T_YREXP4	0.906848
SCH8ORG?	0.863169	BCOL%-LN	0.845726	T_YREXP5	0.912427
ENGABL-N	0.830393	PAR%FARM	0.863609	T_YREXP?	0.986986
ENGABL-?	0.964156	FARM%-LN	0.846970	CERTG-N	0.838771
RD PRI-N	0.841255	PAR%IRRE	0.931918	CERTG-NS	0.873485
RD PRI-?	0.928837	IRRE%-LN	0.824057	CERTG-?	0.836080
%SUBLUN2	0.872255	PAR%WELF	0.951473	T_YREXP2	0.869575
%SUBLUN3	0.880013	WELF%-LN	0.826071	T_YREXP3	0.865145
%SUBLUN4	0.872646	LSTAF-NO	0.917769	T_YREXP4	0.892504
%SUBLUN5	0.871786	LSTAF-PT	0.928304	T_YREXP5	0.900068
%SUBLUN?	0.927956	LSTAF-FT	0.954507	T_YREXP?	0.983698
%REMRED2	0.884657	LSTAF-?	0.981540	UGR ED-Y	0.817813
%REMRED3	0.879905	TCHSUB=1	0.874734	GRA ED-Y	0.851091
%REMRED4	0.824992	TCHSUB-?	0.977676	NOGRAD-Y	0.886197
%REMRED5	0.828156	CMPCL-N?	0.845056	TWKSP-<6	0.889672
%REMRED?	0.929237	CMPLB-N?	0.782912	TWKSP-15	0.924224
%ENREOY2	0.871529	CMPBR-N?	0.825900	TWKSP-35	0.919480
%ENREOY3	0.897894	HS PRI-N	0.834058	TWKSP>35	0.870948

Table C-15 (continued)

Proportion of Variance of the Conditioning Variable Contrasts Accounted for by the Principal Components Used in the Conditioning Model for U.S. History Conditioning Variables, Age 13/Grade 8

Contrast	Proportion of Variance	Contrast	Proportion of Variance	Contrast	Proportion of Variance
TWKSP-?	0.981907	CLASIZ-A	0.911952	RDOTH--M	0.939614
CERT-TMP	0.903859	CLASIZ-L	0.915402	RDOTH--N	0.878373
CERT-REG	0.900304	CLASIZ-M	0.920151	RDOTH--?	0.989282
CERT-HGH	0.927322	CLASIZ-?	0.972265	USEPRM-W	0.885417
CERT-?	0.972853	WK1815-2	0.828667	USEPRM-M	0.939132
TDEG-BAC	0.920749	WK1815-3	0.853758	USEPRM-N	0.899150
TDEG-MAS	0.925744	WK1815-4	0.885461	USEPRM-?	0.989859
TDEG-EDS	0.891700	WK1815-?	0.978833	TESTS--W	0.936755
TDEG-DOC	0.847623	WK1865-2	0.836608	TESTS--M	0.924278
TDEG-PRO	0.975574	WK1865-3	0.883020	TESTS--N	0.841086
TDEG-?	0.983077	WK1865-4	0.867826	TESTS--?	0.989073
PRPER-<1	0.900366	WK1865-?	0.971250	DISCUS-W	0.857704
PRPER-12	0.927986	WK1945-2	0.847727	DISCUS-M	0.898509
PRPER->2	0.947609	WK1945-3	0.852588	DISCUS-?	0.985424
PRPER-?	0.974945	WK1945-4	0.870787	WRTSHT-W	0.848268
CERTH-N	0.883324	WK1945-?	0.977911	WRTSHT-M	0.866261
CERTH-NS	0.855497	WKCURR-2	0.863352	WRTSHT-N	0.836055
CERTH-?	0.899097	WKCURR-3	0.858613	WRTSHT-?	0.989160
CERTG-N	0.856902	WKCURR-4	0.878419	WRTRPT-W	0.844935
CERTG-NS	0.869270	WKCURR-?	0.978385	WRTRPT-M	0.948561
CERTG-?	0.881479	EMPDEM-M	0.788646	WRTRPT-N	0.904025
CERTM-N	0.860271	EMPDEM-L	0.833458	WRTRPT-?	0.982781
CERTM-NS	0.831141	EMPDEM-?	0.985370	GRPPRJ-W	0.914263
CERTM-?	0.877673	EMPCUL-M	0.810172	GRPPRJ-M	0.937736
CERTS-N	0.852863	EMPCUL-L	0.863791	GRPPRJ-N	0.882840
CERTS-NS	0.840808	EMPCUL-?	0.982584	GRPPRJ-?	0.989664
CERTS-?	0.884353	EMPTEC-M	0.827009	ORALRP-W	0.884405
CERTO-N	0.813264	EMPTEC-L	0.849914	ORALRP-M	0.935459
CERTO-NS	0.860843	EMPTEC-?	0.981197	ORALRP-N	0.920124
CERTO-?	0.823228	EMPWRL-M	0.861560	ORALRP-?	0.975427
UGR GE-Y	0.868496	EMPWRL-L	0.866339	USEMAP-W	0.867511
UGR HI-Y	0.815717	EMPWRL-?	0.979762	USEMAP-M	0.887580
UGR SS-Y	0.815131	HISHWK-H	0.905928	USEMAP-N	0.852389
UGR OT-Y	0.833506	HISHWK-1	0.922552	USEMAP-?	0.990148
GRD GE-Y	0.838170	HISHWK-2	0.934278	VIDEOS-W	0.905381
GRD HI-Y	0.852386	HISHW-2+	0.915198	VIDEOS-M	0.958088
GRD SS-Y	0.830307	HISHWK-?	0.979038	VIDEOS-N	0.893040
GRA OT-Y	0.859784	RDTXT--W	0.820852	VIDEOS-?	0.985969
PORT---N	0.822677	RDTXT--M	0.851219	COMPUT-W	0.862462
PORT---?	0.976751	RDTXT--N	0.809779	COMPUT-M	0.907807
CLASAB-N	0.845993	RDTXT--?	0.987390	COMPUT-N	0.958445
CLASAB-?	0.976710	RDOTH--W	0.891030	COMPUT-?	0.985986

Table C-15 (continued)

Proportion of Variance of the Conditioning Variable Contrasts Accounted for by the Principal Components Used in the Conditioning Model for U.S. History Conditioning Variables, Age 13/Grade 8

Contrast	Proportion of Variance	Contrast	Proportion of Variance	Contrast	Proportion of Variance
TRIPS--W	0.896105	MCTEST-?	0.977707	CL826-30	0.877159
TRIPS--M	0.889877	WRTRSP-M	0.828157	CL831-35	0.898816
TRIPS--N	0.949327	WRTRSP-Y	0.883160	CL836+	0.889681
WKLIBR-W	0.860126	WRTRSP-N	0.825950	CL8-MISS	0.842657
WKLIBR-M	0.908846	WRTRSP-?	0.973146	PRUSHI-S	0.835201
WKLIBR-N	0.918535	GRPPRJ-M	0.883023	PRUSHI-U	0.859916
WKLIBR-?	0.978432	GRPPRJ-Y	0.908299	PRUSHI-?	0.971526
GOLIB--W	0.828164	GRPPRJ-N	0.867902	WRKSH8-N	0.825614
GOLIB--M	0.907922	GRPPRJ-?	0.973881	WRKSH8-?	0.925253
GOLIB--N	0.922409	TPORT--M	0.910279	UNICO8-N	0.837559
GOLIB-NO	0.837318	TPORT--Y	0.902726	UNICO8-?	0.904397
GOLIBR-?	0.982245	TPORT--N	0.917264	PROAC8-N	0.811302
LIBOOK-W	0.896643	TPORT--?	0.976376	PROAC8-?	0.904825
LIBOOK-M	0.873569	PEMP-GEO	0.941719	CRL8-DST	0.901993
LIBOOK-N	0.932098	PEMP-SSO	0.870697	CRL8-SCH	0.886432
LIBOK-NO	0.864399	PEMP-SS+	0.861927	CRL8-IND	0.863936
LIBOOK-?	0.958962	PEMP-OTH	0.870933	CRL8-STA	0.878687
COMPAV-D	0.827181	PEMP-?	0.955638	CRL8-EXA	0.845444
COMPAV-C	0.863576	HIST-<50	0.875978	CRL8-?	0.853424
COMPAV-?	0.973542	HIST-<75	0.916621	LKHI8-IN	0.819968
MCTEST-M	0.872542	HIST->75	0.940865	LKHI8-OT	0.868385
MCTEST-Y	0.898183	HIST---?	0.978642	LKHI8-NO	0.896665
MCTEST-N	0.861096	CL821-25	0.889559	LKHI8-?	0.962184

Table C-16

Proportion of Variance of the Conditioning Variable Contrasts Accounted for by the Principal Components Used in the Conditioning Model for U.S. History Conditioning Variables, Age 17/Grade 12

Contrast	Proportion of Variance	Contrast	Proportion of Variance	Contrast	Proportion of Variance
FEMALE	0.929814	HW-YES	0.989575	SEMSCI-^	0.903082
BLACK	0.927562	HWLIN-0	0.978129	#SCI-LIN	0.875557
HISPANIC	0.971341	HWQUAD-0	0.976900	SEMHIS-^	0.894795
ASIAN	0.945022	HITEM=3	0.892545	#HIS-LIN	0.844564
PAC ISLD	0.978980	HITEM=4	0.862252	SEMFLG-^	0.893112
MEXICAN	0.953603	MOMHOM-N	0.975735	#FLG-LIN	0.748816
PUER RIC	0.978099	MOMHOM-?	0.881568	SEMVOC-^	0.886342
CUBN,OTH	0.969188	DADHOM-N	0.936156	#VOC-LIN	0.769070
HISP-?	0.988624	DADHOM-?	0.879193	SEMART-^	0.903862
NON MSA	0.837480	MISS-2<	0.986691	#ART-LIN	0.832940
MID CTY5	0.938900	USA 3-5	0.880262	G/R 22	0.915962
FR/BTWN5	0.932141	USA <3	0.833223	G/R 23	0.920268
SML TWN5	0.947160	USA-?	0.841762	G/R 24	0.964505
RURAL5	0.946900	STGRD1-2	0.956877	G/R 25	0.984124
URBAN FR	0.892686	STGRD3-5	0.951694	G/T 22	0.890236
MED CITY	0.831444	STGRD6>	0.940267	G/T 23	0.857827
SM PLACE	0.896344	SCHCHG-1	0.951457	G/T 24	0.687266
HS GRAD	0.957570	SCHCHG-2	0.970330	G/T 25	0.710274
POST HS	0.949563	SCHCHG-3	0.797540	G/P 22	0.926359
COL GRAD	0.940330	DIS@HOM2	0.898606	G/P 23	0.955858
PARED-?	0.964088	DIS@HOM3	0.916069	G/P 24	0.889277
S EAST	0.886085	DIS@HOM4	0.856681	G/P 25	0.871586
CENTRAL	0.896102	PGS>5	0.815824	G/S 22	0.939222
WEST	0.923142	PGS>10	0.811125	G/S 23	0.936662
PRIVATE	0.918904	COMP-WK	0.928378	R/T 25	0.891089
CATHOLIC	0.916570	COMP-MO	0.920168	R/T 31	0.880116
IEP-NO	0.972014	COMP-NEV	0.907787	R/T 32	0.884283
LEP-NO	0.832781	COMP-?	0.817803	R/T 33	0.892370
CHAP1-N	0.779263	ACADEMIC	0.760660	R/T 34	0.890683
INTEGRAT	0.884553	VOC/TECH	0.905428	R/T 35	0.903156
MINORITY	0.835672	OTHERPGM	0.949914	R/T 41	0.885776
NO LUNCH	0.756627	HS PGM-?	0.802101	R/T 42	0.912986
LUNCH-?	0.860256	VOC/BUSN	0.973585	R/T 43	0.945070
=MA/<MG	0.965711	2 YR COL	0.957157	R/T 44	0.941743
=MA/=MG	0.983292	4 YR COL	0.927525	R/T 45	0.956444
>MA/=MG	0.989203	MILITARY	0.979500	R/T 51	0.950527
HL-SOME	0.875879	OTHERACT	0.981037	R/T 52	0.974833
HL-ALWAY	0.764626	AFTRHS-?	0.770360	R/T 53	0.971774
HL-?	0.853959	SEMENG-^	0.875446	R/T 54	0.971617
TVLIN-0	0.981680	#ENG-LIN	0.891048	R/T 55	0.969082
TV-QUAD	0.981853	SEMMAT-^	0.910786	R/P 25	0.902202
HW-NO	0.988883	#MAT-LIN	0.904909	R/P 31	0.910053

Table C-16 (continued)

Proportion of Variance of the Conditioning Variable Contrasts Accounted for by the Principal Components Used in the Conditioning Model for U.S. History Conditioning Variables, Age 17/Grade 12

Contrast	Proportion of Variance	Contrast	Proportion of Variance	Contrast	Proportion of Variance
R/P 32	0.905810	T/S 43	0.924153	WREPRT-5	0.846513
R/P 33	0.837176	T/S 51	0.941304	WREPRT-?	0.747204
R/P 34	0.896210	T/S 52	0.925727	GRPPRJ-2	0.973531
R/P 35	0.901388	T/S 53	0.949754	GRPPRJ-3	0.974066
R/P 41	0.912987	P/S 32	0.972129	GRPPRJ-4	0.979553
R/P 42	0.838527	P/S 33	0.957254	GRPPRJ-5	0.959635
R/P 43	0.961455	P/S 41	0.960153	GRPPRJ-?	0.785111
R/P 44	0.963414	P/S 42	0.946173	GIVREP-2	0.947049
R/P 45	0.955675	P/S 43	0.950216	GIVREP-3	0.938917
R/P 51	0.903658	P/S 51	0.935821	GIVREP-4	0.980678
R/P 52	0.985530	P/S 52	0.975087	GIVREP-5	0.858946
R/P 53	0.982891	P/S 53	0.970880	GIVREP-?	0.782045
R/P 54	0.981388	PORT--NO	0.955837	MAPSGL-2	0.949060
R/P 55	0.965772	PORT-MIS	0.753042	MAPSGL-3	0.955480
R/S 32	0.973578	RDTEXT-2	0.900726	MAPSGL-4	0.963311
R/S 33	0.971962	RDTEXT-3	0.917821	MAPSGL-5	0.958093
R/S 41	0.974820	RDTEXT-4	0.927231	MAPSGL-?	0.789403
R/S 42	0.958372	RDTEXT-5	0.762718	VIDEO--2	0.960043
R/S 43	0.975097	RDTEXT-?	0.864749	VIDEO--3	0.955704
R/S 51	0.971944	EXTMAT-2	0.948893	VIDEO--4	0.966225
R/S 52	0.983805	EXTMAT-3	0.964411	VIDEO--5	0.885396
R/S 53	0.974414	EXTMAT-4	0.970793	VIDEO--?	0.855653
T/P 25	0.821088	EXTMAT-5	0.954531	COMPUT-2	0.975949
T/P 31	0.790218	EXTMAT-?	0.872349	COMPUT-3	0.975051
T/P 32	0.758897	HISPEO-2	0.981696	COMPUT-4	0.969915
T/P 33	0.896927	HISPEO-3	0.979942	COMPUT-5	0.951830
T/P 34	0.751666	HISPEO-4	0.981988	COMPUT-?	0.816386
T/P 35	0.727337	HISPEO-5	0.972591	QUIZ---2	0.906241
T/P 41	0.725818	HISPEO-?	0.819725	QUIZ---3	0.915200
T/P 42	0.887897	DISMAT-2	0.939316	QUIZ---4	0.983683
T/P 43	0.816712	DISMAT-3	0.930294	QUIZ---5	0.734217
T/P 44	0.805343	DISMAT-4	0.929007	QUIZ---?	0.846578
T/P 45	0.798042	DISMAT-5	0.776836	TRIPS--2	0.986514
T/P 51	0.922939	DISMAT-?	0.774512	TRIPS--3	0.984431
T/P 52	0.814442	WSHANS-2	0.893078	TRIPS--4	0.988423
T/P 53	0.805756	WSHANS-3	0.908394	TRIPS--5	0.982193
T/P 54	0.811527	WSHANS-4	0.964281	TRIPS--?	0.885399
T/P 55	0.920869	WSHANS-5	0.890887	LIBRA--2	0.977146
T/S 32	0.916889	WSHANS-?	0.817006	LIBRA--3	0.984795
T/S 33	0.919732	WREPRT-2	0.948635	LIBRA--4	0.985875
T/S 41	0.920344	WREPRT-3	0.953044	LIBRA--5	0.976208
T/S 42	0.916459	WREPRT-4	0.983213	LIBRA--?	0.898719

BEST COPY AVAILABLE

Table C-16 (continued)

Proportion of Variance of the Conditioning Variable Contrasts Accounted for by the Principal Components Used in the Conditioning Model for U.S. History Conditioning Variables, Age 17/Grade 12

Contrast	Proportion of Variance	Contrast	Proportion of Variance	Contrast	Proportion of Variance
HISGR9-M	0.803938	TESTIMP2	0.962393	RUR%-LIN	0.832718
HISGR9-D	0.803244	TESTIMP3	0.954207	SCH%CITY	0.809398
HISGR9-?	0.826563	TESTIMP4	0.952882	CITY%-LN	0.862750
HISG10-M	0.766793	TESTIMP?	0.965442	PAR%PROF	0.959884
HISG10-D	0.748302	LNGANSW2	0.900633	PROF%-LN	0.850194
HISG10-?	0.814164	LNGANSW3	0.926147	PAR%BCOL	0.957964
HISG11-M	0.919202	LNGANSW4	0.867650	BCOL%-LN	0.849181
HISG11-D	0.810837	LNGANSW?	0.964038	PAR%FARM	0.886024
HISG11-?	0.863959	SNRM-LIN	0.781650	FARM%-LN	0.799509
HISNOW-M	0.813364	ENGABL-N	0.842415	PAR%IRRE	0.920835
HISNOW-?	0.708803	ENGABL-?	0.917847	IRRE%-LN	0.812712
HISHWK-2	0.932218	%SUBLUN2	0.881560	PAR%WELF	0.885882
HISHWK-3	0.932526	%SUBLUN3	0.909674	WELF%-LN	0.815752
HISHWK-4	0.942993	%SUBLUN4	0.911459	LSTAF-NO	0.941692
HISHWK-5	0.929950	%SUBLUN5	0.868413	LSTAF-PT	0.960188
HISHWK-?	0.859736	%SUBLUN?	0.864779	LSTAF-FT	0.965234
PR1815-2	0.875935	%REMRED2	0.856811	LSTAF-?	0.967057
PR1815-3	0.858788	%REMRED3	0.879500	#AP ENG2	0.846331
PR1815-?	0.936681	%REMRED4	0.860010	#AP ENG3	0.850776
181565-2	0.921868	%REMRED5	0.843352	#A? ENG4	0.869916
181565-3	0.848126	%REMRED?	0.904967	#AP ENG5	0.913235
181565-?	0.904160	%ENREOY2	0.898209	#AP ENG6	0.897355
186545-2	0.907453	%ENREOY3	0.910666	#AP ENG?	0.914818
186545-3	0.813338	%ENREOY4	0.871270	#REQENG8	0.849090
186545-?	0.933362	%ENREOY?	0.904326	#REQENG?	0.922876
PS1945-2	0.856609	PARAID-O	0.953889	CMPCL-N?	0.903635
PS1945-3	0.831777	PARAID-N	0.961953	CMPLB-N?	0.785733
PW1945-?	0.901590	PARAID-?	0.978951	CMPBR-N?	0.838768
APHIS-N	0.906153	PARHW-OC	0.900747	HISABL-N	0.851225
APHIS-?	0.792645	PARHW-N?	0.894188	HISABL-?	0.910384
#QUEST+2	0.963165	S/PHW-OC	0.965503	GEOABL-N	0.859017
#QUEST+3	0.958883	S/PHW-N?	0.966665	GEOABL-?	0.847192
#QUEST+4	0.956074	VOLPG-OC	0.897941	#REQ_SS1	0.891681
#QUEST+?	0.878339	VOLPG-N?	0.845025	#REQ_SS2	0.935528
TESTDIF2	0.928057	SCHCH1-N	0.818160	#REQ_SS3	0.880118
TESTDIF3	0.920885	SCHCH1-?	0.935842	#REQ_SS4	0.907047
TESTDIF4	0.936753	%CHAP1-2	0.908974	#REQ_SS5	0.879571
TESTDIF?	0.912993	%CHAP1-3	0.870491	#REQ_SS6	0.939439
TESTEFF2	0.976179	%CHAP1-4	0.895557	#REQ_SS7	0.887434
TESTEFF3	0.961291	%CHAP1-5	0.905549	#REQ_SS8	0.905604
TESTEFF4	0.962763	%CHAP1-?	0.836309	#REQ_SS?	0.918315
TESTEFF?	0.912171	SCH%RURL	0.790143	#REQ_HS1	0.896709

Table C-16 (continued)

Proportion of Variance of the Conditioning Variable Contrasts Accounted for by the Principal Components Used in the Conditioning Model for U.S. History Conditioning Variables, Age 17/Grade 12

Contrast	Proportion of Variance	Contrast	Proportion of Variance	Contrast	Proportion of Variance
#REQ_HS2	0.911998	#REQ_GE1	0.855049	>1SMGE-?	0.900073
#REQ_HS3	0.905877	#REQ_GE2	0.875496	>1SMHS-?	0.956174
#REQ_HS4	0.895737	#REQ_GE3	0.845681	#AP HIS2	0.842679
#REQ_HS5	0.941902	#REQ_GE4	0.818887	#AP HIS3	0.844127
#REQ_HS6	0.915936	#REQ_GE6	0.943958	#AP HIS4	0.903295
#REQ_HS7	0.911268	#REQ_GE8	0.989082	#AP HIS5	0.901743
#REQ_HS8	0.890951	#REQ_GE?	0.831144	#AP HIS6	0.877147
#REQ_HS?	0.848761	>1SMGE-N	0.858819	#AP HIS?	0.917621

Table C-17

Proportion of Variance of the Conditioning Variable Contrasts Accounted for by the Principal Components Used in the Conditioning Model for Geography Conditioning Variables, Age 9/Grade 4

Contrast	Proportion of Variance	Contrast	Proportion of Variance	Contrast	Proportion of Variance
FEMALE	0.926657	HW-NO	0.981974	G/S 23	0.898228
BLACK	0.933675	HW-YES	0.983515	R/T 25	0.902057
HISPANIC	0.969721	HWLIN-0	0.983714	R/T 31	0.914182
ASIAN	0.970060	HWQUAD-0	0.981115	R/T 32	0.892802
PAC ISLD	0.980434	HITEM=3	0.847949	R/T 33	0.916088
MEXICAN	0.960518	HITEM=4	0.807955	R/T 34	0.892875
PUER RIC	0.970566	MOMHOM-N	0.944958	R/T 35	0.887981
CUBN,OTH	0.973933	MOMHOM-?	0.760229	R/T 41	0.875521
HISP-?	0.967008	DADHOM-N	0.919680	R/T 42	0.868354
NON MSA	0.869027	DADHOM-?	0.856365	R/T 43	0.960212
MID CTY5	0.940238	MISS-2<	0.978625	R/T 44	0.959649
FR/BTWN5	0.939897	USA 3-5	0.975865	R/T 45	0.959939
SML TWN5	0.940108	USA <3	0.974530	R/T 52	0.972098
RURAL5	0.949171	USA-?	0.811386	R/T 53	0.967319
URBAN FR	0.875344	STGRD1-2	0.892744	R/T 54	0.956894
MED CITY	0.823361	STGRD3>	0.889524	R/T 55	0.959662
SM PLACE	0.901239	PRESCH-N	0.961100	R/P 25	0.911226
HS GRAD	0.948086	SCHCHG-1	0.902461	R/P 31	0.901969
POST HS	0.928738	SCHCHG-2	0.900247	R/P 32	0.922940
COL GRAD	0.932033	SCHCHG-3	0.848130	R/P 33	0.922390
PARED-?	0.936454	DIS@HOM2	0.871222	R/P 34	0.904059
S EAST	0.875471	DIS@HOM3	0.926967	R/P 35	0.909539
CENTRAL	0.852410	DIS@HOM4	0.817135	R/P 41	0.912936
WEST	0.875995	PGS>5	0.815657	R/P 42	0.907290
PRIVATE	0.923926	PGS>10	0.824650	R/P 43	0.983950
CATHOLIC	0.925430	COMP-WK	0.942100	R/P 44	0.983262
IEP-NO	0.968130	COMP-MO	0.955634	R/P 45	0.977881
LEP-NO	0.860981	COMP-NEV	0.932011	R/P 51	0.976320
CHAP1-N	0.879001	COMP-?	0.796264	R/P 52	0.983738
INTEGRAT	0.853473	G/R 22	0.917466	R/P 53	0.980997
MINORITY	0.823248	G/R 23	0.915436	R/P 54	0.985129
NO LUNCH	0.825416	G/R 24	0.977515	R/P 55	0.983925
LUNCH-?	0.831014	G/R 25	0.984847	R/S 32	0.957082
=MA/<MG	0.873611	G/T 22	0.874760	R/S 33	0.965653
=MA/=MG	0.965067	G/T 23	0.818841	R/S 41	0.966595
=MA/>MG	0.991180	G/T 24	0.721961	R/S 42	0.953214
>MA/=MG	0.963885	G/T 25	0.724865	R/S 43	0.978752
HL-SOME	0.912444	G/P 22	0.941306	R/S 51	0.981633
HL-ALWAY	0.934405	G/P 23	0.919807	R/S 52	0.986080
HL-?	0.819922	G/P 24	0.858533	R/S 53	0.983095
TVLIN-0	0.985309	G/P 25	0.859667	T/P 25	0.838431
TV-QUAD	0.985715	G/S 22	0.918448	T/P 31	0.882842

Table C-17 (continued)

Proportion of Variance of the Conditioning Variable Contrasts Accounted for by the Principal Components Used in the Conditioning Model for Geography Conditioning Variables, Age 9/Grade 4

Contrast	Proportion of Variance	Contrast	Proportion of Variance	Contrast	Proportion of Variance
T/P 32	0.807468	GEOHWK-2	0.956375	NATRES-N	0.901448
T/P 33	0.808127	GEOHWK>2	0.942772	NATRES-?	0.797791
T/P 34	0.776200	GEOHWK-?	0.806804	WEATH--W	0.917220
T/P 35	0.824327	HISPEO-S	0.923786	WEATH--M	0.913214
T/P 41	0.788639	HISPEO-N	0.924134	WEATH--N	0.917832
T/P 42	0.800838	HISPEO-?	0.795737	WEATH--?	0.757733
T/P 43	0.834863	GEOPRJ-S	0.864320	CNTRY--W	0.923832
T/P 44	0.881325	GEOPRJ-N	0.806156	CNTRY--M	0.915273
T/P 45	0.793998	GEOPRJ-?	0.807318	CNTRY--N	0.911235
T/P 51	0.786030	GEOCNT-S	0.889924	CNTRY--?	0.760347
T/P 52	0.833336	GEOCNT-N	0.881990	POPUL--W	0.934051
T/P 53	0.886578	GEOCNT-?	0.774492	POPUL--M	0.935097
T/P 54	0.792768	GEOVID-S	0.885069	POPUL--N	0.922509
T/P 55	0.778612	GEOVID-N	0.870077	POPUL--?	0.821274
T/S 32	0.912184	GEOVID-?	0.872067	CITY---W	0.905115
T/S 33	0.908513	READGE-S	0.874065	CITY---M	0.897071
T/S 41	0.910996	READGE-N	0.857843	CITY---N	0.911396
T/S 42	0.900439	READGE-?	0.874305	CITY---?	0.850714
T/S 43	0.918626	LIBRAY-S	0.893003	ENVIRN-W	0.931232
T/S 51	0.935648	LIBRAY-N	0.859776	ENVIRN-M	0.926716
T/S 52	0.946311	LIBRAY-?	0.860776	ENVIRN-N	0.918206
T/S 53	0.941837	LIKGEO-O	0.871780	ENVIRN-?	0.867236
P/S 32	0.962493	LIKGEO-N	0.821893	REGION-W	0.940473
P/S 33	0.947024	LIKGEO-?	0.837211	REGION-M	0.934659
P/S 41	0.953978	MAPSGL-W	0.934051	REGION-N	0.910832
P/S 42	0.935999	MAPSGL-M	0.936118	REGION-?	0.882771
P/S 43	0.942747	MAPSGL-N	0.922075	#QUEST+2	0.874959
P/S 51	0.929250	MAPSGL-?	0.777591	#QUEST+3	0.872157
P/S 52	0.942533	DISTNS-W	0.951174	#QUEST+4	0.917880
P/S 53	0.929091	DISTNS-M	0.955140	#QUEST+?	0.834162
OFT_3-4W	0.900525	DISTNS-N	0.944410	TESTDIF2	0.897171
OFT_1-2W	0.897350	DISTNS-?	0.797499	TESTDIF3	0.884327
OFT_<1W	0.936487	LATLGS-W	0.947930	TESTDIF4	0.899788
OFT_NEV	0.921526	LATLGS-M	0.940346	TESTDIF?	0.907581
OFT_MIS	0.775781	LATLGS-N	0.924295	TESTEFF2	0.884987
PORT--NO	0.968520	LATLGS-?	0.754378	TESTEFF3	0.870905
PORT-MIS	0.800939	CONTIN-W	0.912308	TESTEFF4	0.910266
COMP-SOM	0.927068	CONTIN-M	0.910035	TESTEFF?	0.889068
COMP--NO	0.923296	CONTIN-N	0.911946	TESTIMP2	0.932764
COMP-MIS	0.833029	CONTIN-?	0.791424	TESTIMP3	0.923624
GEOHWK-H	0.849989	NATRES-W	0.923954	TESTIMP4	0.941644
GEOHWK-1	0.889612	NATRES-M	0.912150	TESTIMP?	0.949136

Table C-17 (continued)

Proportion of Variance of the Conditioning Variable Contrasts Accounted for by the Principal Components Used in the Conditioning Model for Geography Conditioning Variables, Age 9/Grade 4

Contrast	Proportion of Variance	Contrast	Proportion of Variance	Contrast	Proportion of Variance
LNGANSW2	0.883361	SCH%RURL	0.804235	CERTR-NS	0.813880
LNGANSW3	0.891789	RUR%-LIN	0.844549	CERTR-?	0.917085
LNGANSW4	0.865646	SCH%CITY	0.841603	CERTL-N	0.884824
LNGANSW?	0.952381	CITY%-LN	0.883917	CERTL-NS	0.851073
SNRM-LIN	0.816712	PAR%PROF	0.930851	CERTL-?	0.890778
DEPTLIZD	0.853361	PROF%-LN	0.867551	RESOURC2	0.923409
REGROUPD	0.878329	PAR%BCOL	0.942076	RESOURC3	0.918638
SCH4ORG?	0.890081	BCOL%-LN	0.833868	RESOURC4	0.955598
CLAABL-N	0.794909	PAR%FARM	0.860626	RESOURC?	0.986526
CLAABL-?	0.901527	FARM%-LN	0.804829	UGR ED-Y	0.798286
RD PRI-N	0.875647	PAR%IRRE	0.908372	GRA ED-Y	0.871084
RD PRI-?	0.918190	IRRE%-LN	0.828898	NOGRAD-Y	0.880460
%SUBLUN2	0.895533	PAR%WELF	0.918288	T4REDYR2	0.902336
%SUBLUN3	0.911374	WELF%-LN	0.861970	T4REDYR3	0.908790
%SUBLUN4	0.904283	LSTAF-NO	0.900768	T4REDYR4	0.936058
%SUBLUN5	0.881827	LSTAF-PT	0.950497	T4REDYR5	0.951639
%SUBLUN?	0.918954	LSTAF-FT	0.951520	T4REDYR?	0.966726
%REMRED2	0.870899	CMPCL-N?	0.839317	TWKSP-<6	0.867736
%REMRED3	0.888065	CMPLB-N?	0.824857	TWKSP-15	0.852848
%REMRED4	0.853018	CMPBR-N?	0.827291	TWKSP-35	0.890663
%REMRED5	0.852769	HS PRI-N	0.862354	TWKSP>35	0.869806
%REMRED?	0.901079	HS PRI-?	0.912821	TWKSP-?	0.976121
%ENREOY2	0.871988	GE PRI-N	0.845189	CERT-TMP	0.938414
%ENREOY3	0.894252	GE PRI-?	0.922425	CERT-REG	0.948513
%ENREOY4	0.889492	TMCH-PAR	0.843904	CERT-HGH	0.946647
%ENREOY?	0.916328	TMCH-COM	0.990509	CERT-?	0.961363
PARAID-O	0.850646	T_FEMALE	0.816927	TDEG-ASC	0.939429
PARAID-N	0.891295	T_SEX-?	0.986339	TDEG-BAC	0.936122
PARAID-?	0.938048	T_BLACK	0.822705	TDEG-MAS	0.926046
PARHW-OC	0.838211	T_HISP	0.797088	TDEG-EDS	0.917558
PARHW-N?	0.933399	T_ASIAN	0.878106	TDEG-?	0.974501
S/PHW-OC	0.869478	T_PAC IS	0.921448	UGRENG-?	0.850546
S/PHW-N?	0.822891	T_AM IND	0.895039	UGRRED-?	0.830302
VOLPG-OC	0.887398	T_RACE-?	0.961247	GRDENG-?	0.863743
VOLPG-N?	0.816649	T_YREXP2	0.894783	GRDRED-?	0.811920
SCHCH1-N	0.833783	T_YREXP3	0.881663	READDEV2	0.902248
SCHCH1-?	0.960193	T_YREXP4	0.913454	READDEV3	0.911718
%CHAP1-2	0.882896	T_YREXP5	0.930495	READDEV4	0.916941
%CHAP1-3	0.884651	T_YREXP?	0.988522	READDEV5	0.919838
%CHAP1-4	0.854290	CERTG-N	0.880413	READDEV5	0.981777
%CHAP1-5	0.851807	CERTG-?	0.964394	PRPER-<1	0.923317
%CHAP1-?	0.814658	CERTR-N	0.892050	PRPER-12	0.918756

Table C-17 (continued)

Proportion of Variance of the Conditioning Variable Contrasts Accounted for by the Principal Components Used in the Conditioning Model for Geography Conditioning Variables, Age 9/Grade 4

Contrast	Proportion of Variance	Contrast	Proportion of Variance	Contrast	Proportion of Variance
PRPER->2	0.903534	CLASAB-?	0.989448	CITYTC-3	0.911799
PRPER-?	0.969234	CLASIZ-A	0.947775	CITYTC-4	0.881274
CSPECS-N	0.797799	CLASIZ-L	0.931507	CITYTC-?	0.991419
CSPECS-?	0.947209	CLASIZ-M	0.947440	ENVIRT-2	0.909028
CERTH-N	0.923682	CLASIZ-?	0.993573	ENVIRT-3	0.951837
CERTH-NS	0.922744	TIME-45M	0.878818	ENVIRT-4	0.860682
CERTH-?	0.921608	TIME-60M	0.889416	ENVIRT-?	0.988905
CERTG-N	0.932397	TIME-90+	0.877332	REGTCH-2	0.889004
CERTG-NS	0.910665	TIME-MIS	0.985047	REGTCH-3	0.914558
CERTG-?	0.943614	USEMAP-W	0.848738	REGTCH-4	0.872696
CERTO-N	0.814651	USEMAP-M	0.863385	REGTCH-?	0.981546
CERTO-NS	0.882772	USEMAP-N	0.923911	HISHWK-H	0.851260
CERTO-?	0.869662	USEMAP-?	0.987849	HISHWK-1	0.893794
UGR GE-Y	0.877416	DISTNT-W	0.888717	HISHWK-2	0.876557
UGR HI-Y	0.842794	DISTNT-M	0.941217	HISHW-2+	0.902452
UGR SS-Y	0.865745	DISTNT-N	0.882577	HISHWK-?	0.988022
UGR OT-Y	0.850788	DISTNT-?	0.989947	TRIPS--S	0.943589
GRD GE-Y	0.815155	LATLNT-W	0.876463	TRIPS--N	0.948547
GRD HI-Y	0.861913	LATLNT-M	0.937482	TRIPS--?	0.988229
GRD SS-Y	0.785651	LATLNT-N	0.914068	GEOPRJ-S	0.885399
GRA OT-Y	0.889663	LATLNT-?	0.984655	GEOPRJ-N	0.878021
PORT---N	0.813304	CONTCH-W	0.885209	GEOPRJ-?	0.991433
PORT---?	0.979917	CONTCH-M	0.876270	GEOCNT-S	0.941871
PORTR--N	0.790088	CONTCH-N	0.900648	GEOCNT-N	0.954984
PORTR--?	0.976385	CONTCH-?	0.988394	GEOCNT-?	0.987728
T4HISYR2	0.872887	NATRST-2	0.871702	VIDEOT-S	0.879925
T4HISYR3	0.847950	NATRST-3	0.898250	VIDEOT-N	0.889773
T4HISYR4	0.894758	NATRST-4	0.825908	VIDEOT-?	0.993267
T4HISYR5	0.928233	NATRST-?	0.993853	USCOMT-S	0.934005
T4HISYR?	0.974115	CLIMTT-2	0.877776	USCOMT-N	0.944124
TWKSG-<6	0.838658	CLIMTT-3	0.902408	USCOMT-?	0.995316
TWKSG-15	0.875909	CLIMTT-4	0.848136	GOLIB--W	0.932802
TWKSG-35	0.872675	CLIMTT-?	0.995460	GOLIB--M	0.914730
TWKSG>35	0.853183	FORCOU-2	0.903309	GOLIB--N	0.873958
TWKSG-?	0.983714	FORCOU-3	0.931360	GOLIB-NO	0.864683
CSPECG-N	0.938921	FORCOU-4	0.909428	GOLIBR-?	0.992355
CSPECG-?	0.931877	FORCOU-?	0.990088	LIBOOK-W	0.883902
CSPECS-N	0.924935	POPGRO-2	0.881130	LIBOOK-M	0.894044
CSPECS-?	0.931598	POPGRO-3	0.949662	LIBOOK-N	0.867902
CSPECH-N	0.942472	POPGRO-4	0.929390	LIBOK-NO	0.845534
CSPECH-?	0.921821	POPGRO-?	0.985648	LIBOOK-?	0.965953
CLASAB-N	0.895567	CITYTC-2	0.878189	COMPAV-D	0.867915

Table C-17 (continued)

Proportion of Variance of the Conditioning Variable Contrasts Accounted for by the Principal Components Used in the Conditioning Model for Geography Conditioning Variables, Age 9/Grade 4

Contrast	Proportion of Variance	Contrast	Proportion of Variance	Contrast	Proportion of Variance
COMPAV-C	0.883488	CL431-35	0.884083	HISEDU-S	0.875252
COMPAV-?	0.979632	CL436+	0.891300	HISEDU-U	0.875782
MCTEST-M	0.888393	CL4-MISS	0.989042	HISEDU-?	0.947615
MCTEST-Y	0.885878	PRPGEO-A	0.863250	OSSEDU-A	0.849567
MCTEST-N	0.881088	PRPGEO-S	0.853672	OSSEDU-S	0.852558
MCTEST-?	0.980447	PRPGEO-U	0.831402	OSSEDU-U	0.849557
WRTRSP-M	0.859690	PRPGEO-?	0.971422	OSSEDU-?	0.934044
WRTRSP-Y	0.885990	PRPHIS-A	0.872643	SCIEDU-A	0.885145
WRTRSP-N	0.837967	PRPHIS-S	0.887135	SCIEDU-S	0.851960
WRTRSP-?	0.981140	PRPHIS-U	0.836289	SCIEDU-U	0.885313
GRPPRJ-M	0.902231	PRPHIS-?	0.965502	SCIEDU-?	0.912176
GRPPRJ-Y	0.941541	PRPSS--A	0.869860	CRLE-DST	0.876534
GRPPRJ-N	0.860429	PRPSS--S	0.851717	CRLE-SCH	0.878622
GRPPRJ-?	0.980721	PRPSS--U	0.851794	CRLE-IND	0.862772
TPORT--M	0.879858	PRPSS--?	0.965575	CRLE-STA	0.854488
TPORT--Y	0.911251	GEOEDU-A	0.851951	CRLE-EXA	0.897061
TPORT--N	0.916299	GEOEDU-S	0.859423	CRLE-?	0.849500
TPORT--?	0.971261	GEOEDU-U	0.901018	LKGEO-OT	0.810032
CL421-25	0.873688	GEOEDU-?	0.935926	LKGEO-NO	0.895473
CL426-30	0.904653	HISEDU-A	0.819596	LKGEO-?	0.971100

Table C-18

Proportion of Variance of the Conditioning Variable Contrasts Accounted for by the Principal Components Used in the Conditioning Model for Geography Conditioning Variables, Age 13/Grade 8

Contrast	Proportion of Variance	Contrast	Proportion of Variance	Contrast	Proportion of Variance
FEMALE	0.924992	HW-NO	0.973495	G/S 23	0.940261
BLACK	0.921809	HW-YES	0.973727	R/T 25	0.911623
HISPANIC	0.967238	HWLIN-0	0.978217	R/T 31	0.899914
ASIAN	0.964925	HWQUAD-0	0.977388	R/T 32	0.896942
PAC ISLD	0.963799	HITEM=3	0.882402	R/T 33	0.908191
MEXICAN	0.960092	HITEM=4	0.841046	R/T 34	0.901650
PUER RIC	0.980721	MOMHOM-N	0.981926	R/T 35	0.903449
CUBN,OTH	0.979465	MOMHOM-?	0.880617	R/T 41	0.908826
HISP-?	0.984283	DADHOM-N	0.959902	R/T 42	0.910776
NON MSA	0.851481	DADHOM-?	0.907597	R/T 43	0.957260
MID CTY5	0.931283	MISS-2<	0.972461	R/T 44	0.956564
FR/BTWN5	0.931430	USA 3-5	0.967097	R/T 45	0.962448
SML TWN5	0.932019	USA <3	0.939998	R/T 51	0.967500
RURAL5	0.941591	USA-?	0.800899	R/T 52	0.974957
URBAN FR	0.882356	STGRD1-2	0.963616	R/T 53	0.970330
MED CITY	0.835468	STGRD3-5	0.941993	R/T 54	0.963987
SM PLACE	0.899126	STGRD6>	0.914538	R/T 55	0.976088
HS GRAD	0.961189	SCHCHG-1	0.945768	R/P 25	0.908223
POST HS	0.942964	SCHCHG-2	0.949658	R/P 31	0.907486
COL GRAD	0.937573	SCHCHG-3	0.748081	R/P 32	0.913250
PARED-?	0.960498	DIS@HOM2	0.908980	R/P 33	0.863528
S EAST	0.879237	DIS@HOM3	0.935756	R/P 34	0.906329
CENTRAL	0.882679	DIS@HOM4	0.818349	R/P 35	0.907796
WEST	0.865256	PGS>5	0.792819	R/P 41	0.923572
PRIVATE	0.912890	PGS>10	0.799608	R/P 42	0.867705
CATHOLIC	0.913239	COMP-WK	0.955167	R/P 43	0.986854
IEP-NO	0.980362	COMP-MO	0.949164	R/P 44	0.982600
LEP-NO	0.877928	COMP-NEV	0.937633	R/P 45	0.978274
CHAP1-N	0.852800	COMP-?	0.823645	R/P 51	0.961754
INTEGRAT	0.875052	G/R 22	0.912052	R/P 52	0.985607
MINORITY	0.838377	G/R 23	0.908775	R/P 53	0.983889
NO LUNCH	0.774453	G/R 24	0.973762	R/P 54	0.983289
LUNCH-?	0.877259	G/R 25	0.979866	R/P 55	0.975718
=MA/<MG	0.821490	G/T 22	0.822439	R/S 32	0.971795
=MA/=MG	0.948487	G/T 23	0.844507	R/S 33	0.969541
=MA/>MG	0.963456	G/T 24	0.707572	R/S 41	0.975213
>MA/=MG	0.952785	G/T 25	0.714958	R/S 42	0.947197
HL-SOME	0.865551	G/P 22	0.912059	R/S 43	0.980400
HL-ALWAY	0.806016	G/P 23	0.902019	R/S 51	0.980712
HL-?	0.803517	G/P 24	0.847058	R/S 52	0.986071
TVLIN-0	0.985577	G/P 25	0.758877	R/S 53	0.980810
TV-QUAD	0.984984	G/S 22	0.951945	T/P 25	0.807830

Table C-18 (continued)

Proportion of Variance of the Conditioning Variable Contrasts Accounted for by the Principal Components Used in the Conditioning Model for Geography Conditioning Variables, Age 13/Grade 8

Contrast	Proportion of Variance	Contrast	Proportion of Variance	Contrast	Proportion of Variance
T/P 31	0.817207	HISPEO-N	0.966465	WEATH--N	0.923900
T/P 32	0.743534	HISPEO-?	0.875576	WEATH--?	0.731597
T/P 33	0.841246	GEOPRJ-S	0.867823	CNTRY--W	0.937141
T/P 34	0.752663	GEOPRJ-N	0.832780	CNTRY--M	0.912324
T/P 35	0.768163	GEOPRJ-?	0.869811	CNTRY--N	0.886230
T/P 41	0.717777	GEOCNT-S	0.946919	CNTRY--?	0.758049
T/P 42	0.833564	GEOCNT-N	0.945142	POPUL--W	0.926481
T/P 43	0.823325	GEOCNT-?	0.909900	POPUL--M	0.938427
T/P 44	0.840167	GEOVID-S	0.881700	POPUL--N	0.936059
T/P 45	0.801606	GEOVID-N	0.847584	POPUL--?	0.834064
T/P 51	0.881744	GEOVID-?	0.940885	CITY---W	0.932905
T/P 52	0.780287	READGE-S	0.896832	CITY---M	0.920381
T/P 53	0.829433	READGE-N	0.877381	CITY---N	0.901181
T/P 54	0.801020	READGE-?	0.869170	CITY---?	0.855316
T/P 55	0.866838	LIBRAY-S	0.923675	ENVIRN-W	0.952977
T/S 32	0.916806	LIBRAY-N	0.914786	ENVIRN-M	0.945978
T/S 33	0.919228	LIBRAY-?	0.901870	ENVIRN-N	0.936976
T/S 41	0.925398	LIKGEO-O	0.878593	ENVIRN-?	0.857096
T/S 42	0.905283	LIKGEO-N	0.829381	REGION-W	0.945190
T/S 43	0.928832	LIKGEO-?	0.858950	REGION-M	0.944988
T/S 51	0.925951	MAPSGL-W	0.948592	REGION-N	0.931381
T/S 52	0.931225	MAPSGL-M	0.949368	REGION-?	0.869915
T/S 53	0.934907	MAPSGL-N	0.944653	GEOGR6-N	0.785553
P/S 32	0.973052	MAPSGL-?	0.872727	GEOGR6-K	0.806146
P/S 33	0.956469	DISTNS-W	0.958061	GEOGR6-?	0.865816
P/S 41	0.959824	DISTNS-M	0.970411	GEOGR7-N	0.789144
P/S 42	0.948370	DISTNS-N	0.961002	GEOGR7-K	0.789902
P/S 43	0.951689	DISTNS-?	0.860718	GEOGR7-?	0.859065
P/S 51	0.938122	LATLGS-W	0.958714	GEOGR8-N	0.871955
P/S 52	0.973919	LATLGS-M	0.957641	GEOGR8-K	0.880250
P/S 53	0.964849	LATLGS-N	0.947509	GEOGR8-?	0.893617
PORT--NO	0.961329	LATLGS-?	0.813585	EARSCI-N	0.906034
PORT-MIS	0.797855	CONTIN-W	0.939451	EARSCI-?	0.785778
COMP-SOM	0.947989	CONTIN-M	0.924700	#QUEST+2	0.927980
COMP--NO	0.945665	CONTIN-N	0.911636	#QUEST+3	0.924189
COMP-MIS	0.891989	CONTIN-?	0.824655	#QUEST+4	0.941693
GEOHWK-H	0.882575	NATRES-W	0.942942	#QUEST+?	0.845318
GEOHWK-1	0.903467	NATRES-M	0.949961	TESTDIF2	0.934240
GEOHWK-2	0.938865	NATRES-N	0.937924	TESTDIF3	0.930698
GEOHWK>2	0.947000	NATRES-?	0.823404	TESTDIF4	0.936350
GEOHWK-?	0.858232	WEATH--W	0.940111	TESTDIF?	0.838645
HISPEO-S	0.965757	WEATH--M	0.938570	TESTEFF2	0.927627

Table C-18 (continued)

Proportion of Variance of the Conditioning Variable Contrasts Accounted for by the Principal Components Used in the Conditioning Model for Geography Conditioning Variables, Age 13/Grade 8

Contrast	Proportion of Variance	Contrast	Proportion of Variance	Contrast	Proportion of Variance
TESTEFF3	0.912263	SCHCH1-N	0.812680	GEOABL-N	0.903390
TESTEFF4	0.934847	SCHCH1-?	0.958428	GEOABL-?	0.846181
TESTEFF?	0.870791	%CHAP1-2	0.872844	TMCH-PAR	0.930300
TESTIMP2	0.921395	%CHAP1-3	0.870566	TMCH-COM	0.971140
TESTIMP3	0.909972	%CHAP1-4	0.850842	T_FEMALE	0.778586
TESTIMP4	0.914986	%CHAP1-5	0.849090	T_SEX-?	0.985511
TESTIMP?	0.906926	%CHAP1-?	0.791633	T_BLACK	0.888736
LNGANSW2	0.886934	SCH%RURL	0.785076	T_HISP	0.931944
LNGANSW3	0.896356	RUR%-LIN	0.862552	T_ASIAN	0.840588
LNGANSW4	0.867996	SCH%CITY	0.800641	T_PAC IS	0.931320
LNGANSW?	0.905468	CITY%-LN	0.862768	T_RACE-?	0.979046
SNRM-LIN	0.856492	PAR%PROF	0.958133	T_YREXP2	0.866846
SEMIDEPT	0.913561	PROF%-LN	0.862551	T_YREXP3	0.880628
DEPTLIZD	0.922803	PAR%BCOL	0.958631	T_YREXP4	0.908938
SCH8ORG?	0.873788	BCOL%-LN	0.866719	T_YREXP5	0.920091
ENGABL-N	0.822972	PAR%FARM	0.864195	T_YREXP?	0.990897
ENGABL-?	0.962164	FARM%-LN	0.835278	CERTG-N	0.825232
RD PRI-N	0.834166	PAR%IRRE	0.932353	CERTG-NS	0.906880
RD PRI-?	0.924563	IRRE%-LN	0.799233	CERTG-?	0.846046
%SUBLUN2	0.866707	PAR%WELF	0.947482	T_YREXP2	0.858995
%SUBLUN3	0.877297	WELF%-LN	0.825063	T_YREXP3	0.869436
%SUBLUN4	0.892507	LSTAF-NO	0.910135	T_YREXP4	0.891746
%SUBLUN5	0.867555	LSTAF-PT	0.926798	T_YREXP5	0.912092
%SUBLUN?	0.922458	LSTAF-FT	0.948745	T_YREXP?	0.990338
%REMRED2	0.883214	LSTAF-?	0.984739	UGR ED-Y	0.809925
%REMRED3	0.881025	TCHSUB=1	0.871975	GRA ED-Y	0.841528
%REMRED4	0.837018	TCHSUB-?	0.972092	NOGRAD-Y	0.879970
%REMRED5	0.853423	CMPCL-N?	0.847059	TWKSP-<6	0.911793
%REMRED?	0.925732	CMPLB-N?	0.797251	TWKSP-15	0.906915
%ENREOY2	0.898721	CMPBR-N?	0.854549	TWKSP-35	0.907761
%ENREOY3	0.893360	HS PRI-N	0.821272	TWKSP>35	0.922936
%ENREOY4	0.881512	HS PRI-?	0.911058	TWKSP-?	0.986715
%ENREOY?	0.919251	GE PRI-N	0.883229	CERT-TMP	0.912528
PARAID-O	0.879023	GE PRI-?	0.908134	CERT-REG	0.929302
PARAID-N	0.912886	HTSUB#=1	0.902404	CERT-HGH	0.937288
PARAID-?	0.979023	NO HIST8	0.850814	CERT-?	0.972150
PARHW-OC	0.869185	HTSUB#-?	0.962793	TDEG-BAC	0.925022
PARHW-N?	0.844311	GTSUB#=1	0.928106	TDEG-MAS	0.917726
S/PHW-OC	0.920300	NO GEOG8	0.917063	TDEG-EDS	0.884956
S/PHW-N?	0.930495	GTSUB#-?	0.898238	TDEG-DOC	0.799843
VOLPG-OC	0.865633	HISABL-N	0.893670	TDEG-PRO	0.967657
VOLPG-N?	0.828152	HISABL-?	0.929917	TDEG-?	0.988566

Table C-18 (continued)

Proportion of Variance of the Conditioning Variable Contrasts Accounted for by the Principal Components Used in the Conditioning Model for Geography Conditioning Variables, Age 13/Grade 8

Contrast	Proportion of Variance	Contrast	Proportion of Variance	Contrast	Proportion of Variance
PRPER-<1	0.924410	DISTNT-?	0.990609	TRIPS--S	0.909224
PRPER-12	0.929255	LATLNT-W	0.899394	TRIPS--N	0.954351
PRPER->2	0.954082	LATLNT-M	0.937355	TRIPS--?	0.991883
PRPER-?	0.979592	LATLNT-N	0.910342	GEOPRJ-S	0.875978
CERTH-N	0.878341	LATLNT-?	0.993409	GEOPRJ-N	0.875479
CERTH-NS	0.836197	CONTCH-W	0.868843	GEOPRJ-?	0.993149
CERTH-?	0.884487	CONTCH-M	0.877660	GEOCNT-S	0.938152
CERTG-N	0.863229	CONTCH-N	0.826756	GEOCNT-N	0.943689
CERTG-NS	0.872486	CONTCH-?	0.988643	GEOCNT-?	0.987988
CERTG-?	0.872290	NATRST-2	0.861417	VIDEOT-S	0.877433
CERTM-N	0.812447	NATRST-3	0.901720	VIDEOT-N	0.881691
CERTM-NS	0.829002	NATRST-4	0.872498	VIDEOT-?	0.990296
CERTM-?	0.881423	NATRST-?	0.991454	USCOMT-S	0.914000
CERTS-N	0.848400	CLIMTT-2	0.871788	USCOMT-N	0.946511
CERTS-NS	0.856641	CLIMTT-3	0.900752	USCOMT-?	0.990487
CERTS-?	0.871609	CLIMTT-4	0.880228	GOLIB--W	0.848208
CERTO-N	0.813479	CLIMTT-?	0.990038	GOLIB--M	0.935629
CERTO-NS	0.877137	FORCOU-2	0.889478	GOLIB--N	0.928946
CERTO-?	0.832777	FORCOU-3	0.876933	GOLIB-NO	0.841424
UGR GE-Y	0.876705	FORCOU-4	0.862230	GOLIBR-?	0.987293
UGR HI-Y	0.805605	FORCOU-?	0.987616	LIBOOK-W	0.884053
UGR SS-Y	0.801311	POPGRO-2	0.883715	LIBOOK-M	0.917102
UGR OT-Y	0.825965	POPGRO-3	0.898705	LIBOOK-N	0.950657
GRD GE-Y	0.867037	POPGRO-4	0.886129	LIBOK-NO	0.874526
GRD HI-Y	0.861862	POPGRO-?	0.995487	LIBOOK-?	0.975574
GRD SS-Y	0.828281	CITYTC-2	0.867142	COMPAV-D	0.878161
GRA OT-Y	0.880779	CITYTC-3	0.880062	COMPAV-C	0.882985
PORT---N	0.813534	CITYTC-4	0.840105	COMPAV-?	0.978150
PORT---?	0.977527	CITYTC-?	0.987665	MCTEST-M	0.853823
CLASAB-N	0.871341	ENVIRT-2	0.860930	MCTEST-Y	0.848950
CLASAB-?	0.972885	ENVIRT-3	0.927516	MCTEST-N	0.843646
CLASIZ-A	0.919648	ENVIRT-4	0.907081	MCTEST-?	0.979752
CLASIZ-L	0.916642	ENVIRT-?	0.991156	WRTRSP-M	0.828962
CLASIZ-M	0.923198	REGTCH-2	0.879388	WRTRSP-Y	0.856411
CLASIZ-?	0.970904	REGTCH-3	0.921940	WRTRSP-N	0.824165
USEMAP-W	0.852069	REGTCH-4	0.857725	WRTRSP-?	0.972167
USEMAP-M	0.888629	REGTCH-?	0.989767	GRPPRJ-M	0.884523
USEMAP-N	0.838215	HISHWK-H	0.867555	GRPPRJ-Y	0.915024
USEMAP-?	0.992553	HISHWK-1	0.903161	GRPPRJ-N	0.892925
DISTNT-W	0.874958	HISHWK-2	0.878554	GRPPRJ-?	0.986590
DISTNT-M	0.924935	HISHW-2+	0.954485	TPORT--M	0.914957
DISTNT-N	0.911477	HISHWK-?	0.990051	TPORT--Y	0.884610

Table C-18 (continued)

Proportion of Variance of the Conditioning Variable Contrasts Accounted for by the Principal Components Used in the Conditioning Model for Geography Conditioning Variables, Age 13/Grade 8

Contrast	Proportion of Variance	Contrast	Proportion of Variance	Contrast	Proportion of Variance
TPORT--N	0.929179	HISEDU-A	0.854951	PRPGEO-U	0.855927
TPORT--?	0.982249	HISEDU-S	0.851221	PRPGEO-?	0.963886
PEMP-GEO	0.890281	HISEDU-U	0.861937	PRPHIS-A	0.868761
PEMP-SSO	0.886211	HISEDU-?	0.946679	PRPHIS-S	0.878521
PEMP-OTH	0.844860	OSSEDU-A	0.854825	PRPHIS-U	0.878968
PEMP-?	0.973827	OSSEDU-S	0.857224	PRPHIS-?	0.949153
CL821-25	0.876028	OSSEDU-U	0.863462	PRPOSS-A	0.836968
CL826-30	0.881956	OSSEDU-?	0.924476	PRPOSS-S	0.888268
CL831-35	0.884676	CRL8-DST	0.886181	PRPOSS-U	0.852490
CL836+	0.884708	CRL8-SCH	0.904028	PRPOSS-?	0.934208
CL8-MISS	0.874753	CRL8-IND	0.878392	LKGE8-IN	0.829414
GEOEDU-A	0.852397	CRL8-EXA	0.920451	LKGE8-OT	0.877913
GEOEDU-S	0.854223	CRL8-?	0.894948	LKGE8-NO	0.876345
GEOEDU-U	0.870967	PRPGEO-A	0.840663	LKGE8-?	0.940445
GEOEDU-?	0.932572	PRPGEO-S	0.876462		

Table C-19

Proportion of Variance of the Conditioning Variable Contrasts Accounted for by the Principal Components Used in the Conditioning Model for Geography Conditioning Variables, Age 17/Grade 12

Contrast	Proportion of Variance	Contrast	Proportion of Variance	Contrast	Proportion of Variance
FEMALE	0.930032	HW-YES	0.987599	SEMSCI-^	0.903472
BLACK	0.928089	HWLIN-0	0.975874	#SCI-LIN	0.877455
HISPANIC	0.973025	HWQUAD-0	0.973467	SEMHIS-^	0.901203
ASIAN	0.950220	HITEM=3	0.889389	#HIS-LIN	0.845326
PAC ISLD	0.975366	HITEM=4	0.858707	SEMFLG-^	0.890715
MEXICAN	0.948806	MOMHOM-N	0.922171	#FLG-LIN	0.750340
PUER RIC	0.972127	MOMHOM-?	0.866421	SEMVOC-^	0.889260
CUBN,OTH	0.969649	DADHOM-N	0.873991	#VOC-LIN	0.802358
HISP-?	0.982454	DADHOM-?	0.891544	SEMART-^	0.906484
NON MSA	0.832598	MISS-2<	0.980907	#ART-LIN	0.805117
MID CTY5	0.940019	USA 3-5	0.896178	G/R 22	0.915835
FR/BTWN5	0.935808	USA <3	0.829481	G/R 23	0.923901
SML TWN5	0.949906	USA-?	0.820796	G/R 24	0.968147
RURAL5	0.947726	STGRD1-2	0.938305	G/R 25	0.982955
URBAN FR	0.895323	STGRD3-5	0.948647	G/T 22	0.873545
MED CITY	0.803669	STGRD6>	0.929032	G/T 23	0.872605
SM PLACE	0.903955	SCHCHG-1	0.959888	G/T 24	0.710277
HS GRAD	0.958742	SCHCHG-2	0.969878	G/T 25	0.725155
POST HS	0.950874	SCHCHG-3	0.848767	G/P 22	0.915483
COL GRAD	0.941652	DIS@HOM2	0.909321	G/P 23	0.948163
PARED-?	0.961582	DIS@HOM3	0.924160	G/P 24	0.893592
S EAST	0.895089	DIS@HOM4	0.861159	G/P 25	0.878764
CENTRAL	0.901391	PGS>5	0.813553	G/S 22	0.941225
WEST	0.917954	PGS>10	0.826094	G/S 23	0.938141
PRIVATE	0.930210	COMP-WK	0.934059	R/T 25	0.897559
CATHOLIC	0.924218	COMP-MO	0.921156	R/T 31	0.888156
IEP-NO	0.976097	COMP-NEV	0.903052	R/T 32	0.883318
LEP-NO	0.852304	COMP-?	0.821629	R/T 33	0.889742
CHAP1-N	0.854667	ACADEMIC	0.745701	R/T 34	0.894603
INTEGRAT	0.896182	VOC/TECH	0.918157	R/T 35	0.909364
MINORITY	0.829859	OTHERPGM	0.947410	R/T 41	0.899873
NO LUNCH	0.780835	HS PGM-?	0.811638	R/T 42	0.922807
LUNCH-?	0.865207	VOC/BUSN	0.972395	R/T 43	0.949739
=MA/<MG	0.960595	2 YR COL	0.957785	R/T 44	0.943620
=MA/=MG	0.982005	4 YR COL	0.934796	R/T 45	0.956426
>MA/=MG	0.986142	MILITARY	0.979261	R/T 51	0.956105
HL-SOME	0.869485	OTHERACT	0.980610	R/T 52	0.977163
HL-ALWAY	0.775588	AFTRHS-?	0.786949	R/T 53	0.970067
HL-?	0.831174	SEMENG-^	0.878846	R/T 54	0.972509
TVLIN-0	0.982293	#ENG-LIN	0.890768	R/T 55	0.960656
TV-QUAD	0.982058	SEMMAT-^	0.916891	R/P 25	0.911732
HW-NO	0.987559	#MAT-LIN	0.903603	R/P 31	0.911745

Table C-19 (continued)

Proportion of Variance of the Conditioning Variable Contrasts Accounted for by the Principal Components Used in the Conditioning Model for Geography Conditioning Variables, Age 17/Grade 12

Contrast	Proportion of Variance	Contrast	Proportion of Variance	Contrast	Proportion of Variance
R/P 32	0.906177	T/S 43	0.923205	LIKGEO-?	0.674675
R/P 33	0.828185	T/S 51	0.939100	MAPSGL-W	0.906062
R/P 34	0.901378	T/S 52	0.922993	MAPSGL-M	0.965157
R/P 35	0.902178	T/S 53	0.949327	MAPSGL-N	0.924378
R/P 41	0.909266	P/S 32	0.973839	MAPSGL-?	0.885484
R/P 42	0.842907	P/S 33	0.956978	DISTNS-W	0.822427
R/P 43	0.965605	P/S 41	0.962331	DISTNS-M	0.980048
R/P 44	0.970302	P/S 42	0.947056	DISTNS-N	0.910441
R/P 45	0.961182	P/S 43	0.954930	DISTNS-?	0.897153
R/P 51	0.909112	P/S 51	0.934890	LATLGS-W	0.882750
R/P 52	0.984665	P/S 52	0.981826	LATLGS-M	0.971996
R/P 53	0.983072	P/S 53	0.973005	LATLGS-N	0.929661
R/P 54	0.981805	PORT--NO	0.965413	LATLGS-?	0.820952
R/P 55	0.966282	PORT-MIS	0.824244	CONTIN-W	0.898021
R/S 32	0.973011	COMP-SOM	0.968368	CONTIN-M	0.950524
R/S 33	0.973920	COMP--NO	0.967946	CONTIN-N	0.887774
R/S 41	0.975740	COMP-MIS	0.946880	CONTIN-?	0.866022
R/S 42	0.957976	GEOHWK-H	0.886846	NATRES-W	0.880552
R/S 43	0.977045	GEOHWK-1	0.912984	NATRES-M	0.962212
R/S 51	0.978820	GEOHWK-2	0.940763	NATRES-N	0.907967
R/S 52	0.987768	GEOHWK>2	0.952109	NATRES-?	0.850222
R/S 53	0.976338	GEOHWK-?	0.930153	WEATH--W	0.909885
T/P 25	0.816819	HISPEO-S	0.974737	WEATH--M	0.951067
T/P 31	0.791946	HISPEO-N	0.972334	WEATH--N	0.899285
T/P 32	0.771123	HISPEO-?	0.952206	WEATH--?	0.799782
T/P 33	0.902062	GEOPRJ-S	0.933338	CNTRY--W	0.871228
T/P 34	0.745095	GEOPRJ-N	0.923745	CNTRY--M	0.925505
T/P 35	0.735809	GEOPRJ-?	0.939000	CNTRY--N	0.817499
T/P 41	0.732323	GEOCNT-S	0.963659	CNTRY--?	0.816527
T/P 42	0.894482	GEOCNT-N	0.957880	POPUL--W	0.864354
T/P 43	0.816379	GEOCNT-?	0.950648	POPUL--M	0.946444
T/P 44	0.806802	GEOVID-S	0.934794	POPUL--N	0.853876
T/P 45	0.801733	GEOVID-N	0.919067	POPUL--?	0.848702
T/P 51	0.936330	GEOVID-?	0.909397	CITY---W	0.879372
T/P 52	0.807620	READGE-S	0.915996	CITY---M	0.932024
T/P 53	0.804999	READGE-N	0.902933	CITY---N	0.836787
T/P 54	0.819371	READGE-?	0.956177	CITY---?	0.848830
T/P 55	0.926122	LIBRAY-S	0.961831	ENVIRN-W	0.954012
T/S 32	0.916061	LIBRAY-N	0.964261	ENVIRN-M	0.951369
T/S 33	0.921411	LIBRAY-?	0.938123	ENVIRN-N	0.928028
T/S 41	0.918339	LIKGEO-O	0.895110	ENVIRN-?	0.862164
T/S 42	0.912841	LIKGEO-N	0.864223	REGION-W	0.942218

Table C-19 (continued)

Proportion of Variance of the Conditioning Variable Contrasts Accounted for by the Principal Components Used in the Conditioning Model for Geography Conditioning Variables, Age 17/Grade 12

Contrast	Proportion of Variance	Contrast	Proportion of Variance	Contrast	Proportion of Variance
REGION-M	0.961470	LNGANSW3	0.933695	PAR%BCOL	0.958801
REGION-N	0.947327	LNGANSW4	0.885402	BCOL%-LN	0.851473
REGION-?	0.876913	LNGANSW?	0.966047	PAR%FARM	0.886139
DIFFUS-W	0.959759	SNRM-LIN	0.791601	FARM%-LN	0.803079
DIFFUS-M	0.955866	ENGABL-N	0.846225	PAR%IRRE	0.924169
DIFFUS-N	0.945500	ENGABL-?	0.922288	IRRE%-LN	0.796229
DIFFUS-?	0.908959	%SUBLUN2	0.883677	PAR%WELF	0.893424
GEOGR9-N	0.753543	%SUBLUN3	0.909942	WELF%-LN	0.836882
GEOGR9-K	0.778996	%SUBLUN4	0.906631	LSTAF-NO	0.952142
GEOGR9-?	0.814681	%SUBLUN5	0.863035	LSTAF-PT	0.959510
GEOG10-N	0.730486	%SUBLUN?	0.862285	LSTAF-FT	0.966239
GEOG10-K	0.793157	%REMRED2	0.856361	LSTAF-?	0.966761
GEOG10-?	0.839562	%REMRED3	0.882271	#AP ENG2	0.846905
GEOG11-N	0.764492	%REMRED4	0.860116	#AP ENG3	0.843872
GEOG11-K	0.738159	%REMRED5	0.858179	#AP ENG4	0.871127
GEOG11-?	0.808593	%REMRED?	0.908434	#AP ENG5	0.921974
GEOG12-N	0.838895	%ENREOY2	0.886735	#AP ENG6	0.879693
GEOG12-K	0.901611	%ENREOY3	0.907831	#AP ENG?	0.922351
GEOG12-?	0.868629	%ENREOY4	0.875079	#REQENG8	0.867721
WLDGEO-N	0.722920	%ENREOY?	0.910580	#REQENG?	0.913488
WLDGEO-?	0.856095	PARAID-O	0.957051	CMPCL-N?	0.884308
USAGEO-N	0.830059	PARAID-N	0.961560	CMPLB-N?	0.781782
USAGEO-?	0.810926	PARAID-?	0.978739	CMPBR-N?	0.897398
ADVPLM-N	0.949264	PARHW-OC	0.906463	HISABL-N	0.852678
ADVPLM-?	0.804050	PARHW-N?	0.900684	HISABL-?	0.918445
#QUEST+2	0.945124	S/PHW-OC	0.958860	GEOABL-N	0.862177
#QUEST+3	0.937596	S/PHW-N?	0.963775	GEOABL-?	0.851232
#QUEST+4	0.940705	VOLPG-OC	0.898950	#REQ_SS1	0.885493
#QUEST+?	0.832267	VOLPG-N?	0.841947	#REQ_SS2	0.910830
TESTDIF2	0.943390	SCHCH1-N	0.819950	#REQ_SS3	0.926342
TESTDIF3	0.930580	SCHCH1-?	0.934626	#REQ_SS4	0.890257
TESTDIF4	0.940576	%CHAP1-2	0.907269	#REQ_SS5	0.924479
TESTDIF?	0.917576	%CHAP1-3	0.884193	#REQ_SS6	0.871695
TESTEFF2	0.972386	%CHAP1-4	0.889351	#REQ_SS7	0.945296
TESTEFF3	0.956042	%CHAP1-5	0.939879	#REQ_SS8	0.871681
TESTEFF4	0.956180	%CHAP1-?	0.819282	#REQ_HS1	0.874644
TESTEFF?	0.916046	SCH%RURL	0.823956	#REQ_HS2	0.900130
TESTIMP2	0.957054	RUR%-LIN	0.822713	#REQ_HS3	0.895916
TESTIMP3	0.956542	SCH%CITY	0.812231	#REQ_HS4	0.902603
TESTIMP4	0.952453	CITY%-LN	0.866646	#REQ_HS5	0.913381
TESTIMP?	0.966590	PAR%PROF	0.961752	#REQ_HS6	0.916315
LNGANSW2	0.931129	PROF%-LN	0.840940	#REQ_HS7	0.887749

Table C-19 (continued)

Proportion of Variance of the Conditioning Variable Contrasts Accounted for by the Principal Components Used in the Conditioning Model for Geography Conditioning Variables, Age 17/Grade 12

Contrast	Proportion of Variance	Contrast	Proportion of Variance	Contrast	Proportion of Variance
#REQ_HS8	0.911876	#REQ_GE7	0.929964	#AP HIS4	0.891830
#REQ_GE1	0.921709	>1SMGE-N	0.839293	#AP HIS5	0.898415
#REQ_GE2	0.893442	>1SMGE-?	0.905502	#AP HIS6	0.874155
#REQ_GE3	0.915412	>1SMHS-?	0.954378	#AP HIS?	0.923257
#REQ_GE4	0.856820	#AP HIS2	0.840805		
#REQ_GE5	0.828967	#AP HIS3	0.853034		

Appendix D

IRT PARAMETERS

Appendix D

IRT PARAMETERS

This appendix contains tables of IRT (item response theory) parameters for NAEP items that were scaled in each subject area and study (main and long-term trend) for which IRT scales were created.

For each of the binary scored items used in scaling (i.e., multiple-choice items and short constructed-response items), the tables provide estimates of the IRT parameters (which correspond to a_j, b_j, and c_j in Equation 11.1 in Chapter 11) and their associated standard errors (s.e.) of the estimates. For each of the polytomously scored items (i.e., the extended constructed-response items), the tables also show the estimates of the d_{jv} parameters (see Equation 11.3) and their associated standard errors.

The tables also show the block in which each item appears for each age class (*Block*) and the position of each item within its block (*Item*).

Note that item parameters shown in this appendix are in the metrics used for the original calibration of the scales. The transformations needed to represent these parameters in terms of the metric of the final reporting scales are given in Chapters 12 through 18.

Table D-1

IRT Parameters for the Reading Main Samples
Reading for Literary Experience Scale, Age 9/Grade 4

NAEP ID	Block	Item	A	S.E.	B	S.E.	C	S.E.	D	S.E.
R012001	RC	1	1.786	(0.087)	0.683	(0.018)	0.102	(0.009)		
R012002	RC	2	1.668	(0.050)	0.071	(0.014)	0.000	(0.000)		
R012003	RC	3	1.871	(0.085)	-0.348	(0.029)	0.201	(0.017)		
R012004	RC	4	0.822	(0.031)	0.564	(0.026)	0.000	(0.000)		
R012005	RC	5	1.325	(0.072)	0.292	(0.032)	0.208	(0.015)		
R012006	RC	6	0.575	(0.018)	0.941	(0.023)	0.000	(0.000)	0.175	(0.041)
									0.058	(0.052)
									-0.233	(0.063)
R012007	RC	7	0.883	(0.056)	-0.341	(0.074)	0.251	(0.028)		
R012008a*	RC	8	0.634	(0.035)	-0.295	(0.050)	0.000	(0.000)		
R012008b*	RC	8	1.067	(0.053)	-0.393	(0.037)	0.000	(0.000)		
R012009	RC	9	1.589	(0.095)	-0.458	(0.050)	0.335	(0.024)		
R012010a*	RC	10	1.403	(0.066)	-0.045	(0.028)	0.000	(0.000)		
R012010b*	RC	10	1.281	(0.064)	-0.348	(0.033)	0.000	(0.000)		
R012011	RC	11	2.018	(0.117)	-0.041	(0.031)	0.256	(0.018)		
R012101	RD	1	2.073	(0.111)	-0.667	(0.035)	0.319	(0.021)		
R012102a*	RD	2	0.941	(0.044)	0.167	(0.030)	0.000	(0.000)		
R012102b*	RD	2	0.894	(0.042)	-0.928	(0.048)	0.000	(0.000)		
R012103	RD	3	1.445	(0.066)	-0.284	(0.035)	0.195	(0.018)		
R012104	RD	4	0.795	(0.028)	-0.107	(0.026)	0.000	(0.000)		
R012105	RD	5	0.898	(0.054)	0.126	(0.052)	0.180	(0.021)		
R012106a*	RD	6	1.140	(0.052)	0.222	(0.027)	0.000	(0.000)		
R012106b*	RD	6	0.950	(0.049)	0.523	(0.032)	0.000	(0.000)		
R012107	RD	7	1.298	(0.079)	0.316	(0.037)	0.245	(0.017)		
R012108	RD	8	0.720	(0.027)	-0.925	(0.043)	0.000	(0.000)		
R012109	RD	9	0.649	(0.026)	-0.892	(0.047)	0.000	(0.000)		
R012110	RD	10	0.899	(0.059)	-0.824	(0.099)	0.302	(0.036)		
R012111	RD	11	1.083	(0.037)	1.455	(0.023)	0.000	(0.000)	0.843	(0.022)
									-0.843	(0.060)
R012112	RD	12	0.869	(0.037)	-0.473	(0.037)	0.000	(0.000)		
R012401b**	RI	1	1.027	(0.044)	1.861	(0.031)	0.000	(0.000)	1.267	(0.029)
									-0.421	(0.079)
									-0.845	(0.303)
R012402b**	RI	2	1.019	(0.101)	0.071	(0.087)	0.333	(0.033)		
R012403b**	RI	3	1.221	(0.066)	0.940	(0.034)	0.000	(0.000)		
R012404b**	RI	4	1.164	(0.097)	0.270	(0.055)	0.217	(0.025)		
R012405b**	RI	5	1.362	(0.144)	0.900	(0.046)	0.211	(0.019)		
R012406b**	RI	6	0.977	(0.051)	0.460	(0.032)	0.000	(0.000)		
R012407b**	RI	7	1.071	(0.052)	-0.188	(0.033)	0.000	(0.000)		
R012408b**	RI	8	1.608	(0.145)	0.303	(0.048)	0.287	(0.024)		
R012409b**	RI	9	1.462	(0.080)	0.702	(0.028)	0.000	(0.000)		
R012601	RE	1	0.905	(0.039)	1.236	(0.039)	0.000	(0.000)		

__Note__: a = item parameters are based on only 1994 data; b = item parameters are based on only 1992 data.

**__Note__: This block name identifies two different sets of items for 1992 and 1994–a=1994 items; b=1992 items.*

Table D-1 (continued)

IRT Parameters for the Reading Main Samples
Reading for Literary Experience Scale, Age 9/Grade 4

NAEP ID	Block	Item	A	S.E.	B	S.E.	C	S.E.	D	S.E.
R012602	RE	2	1.931	(0.099)	1.341	(0.031)	0.193	(0.007)		
R012603	RE	3	1.623	(0.085)	0.269	(0.027)	0.211	(0.015)		
R012604	RE	4	1.197	(0.050)	1.196	(0.030)	0.000	(0.000)		
R012605	RE	5	1.220	(0.116)	1.038	(0.043)	0.312	(0.015)		
R012606	RE	6	1.520	(0.095)	0.476	(0.031)	0.270	(0.015)		
R012607	RE	7	1.135	(0.036)	1.394	(0.018)	0.000	(0.000)	0.639	(0.022)
									0.201	(0.031)
									-0.840	(0.074)
R012608	RE	8	0.900	(0.071)	-0.177	(0.091)	0.373	(0.031)		
R012609	RE	9	1.055	(0.092)	0.851	(0.045)	0.230	(0.018)		
R012610	RE	10	2.405	(0.145)	0.747	(0.025)	0.395	(0.012)		
R012611a*	RE	11	0.874	(0.050)	0.408	(0.038)	0.000	(0.000)		
R012611b*	RE	11	0.855	(0.049)	-0.129	(0.041)	0.000	(0.000)		
R015801a**	RI	1	1.109	(0.077)	-1.040	(0.086)	0.228	(0.035)		
R015802a**	RI	2	0.563	(0.031)	-0.556	(0.056)	0.000	(0.000)		
R015803a**	RI	3	0.645	(0.021)	0.064	(0.035)	0.000	(0.000)	1.438	(0.056)
									-1.438	(0.052)
R015804a**	RI	4	0.731	(0.024)	0.919	(0.029)	0.000	(0.000)	1.948	(0.050)
									-0.255	(0.046)
									-1.693	(0.116)
R015805a**	RI	5	1.055	(0.111)	0.472	(0.068)	0.317	(0.026)		
R015806a**	RI	6	0.681	(0.026)	0.452	(0.035)	0.000	(0.000)	1.132	(0.050)
									-1.132	(0.057)
R015807a**	RI	7	0.680	(0.029)	0.077	(0.034)	0.000	(0.000)	0.992	(0.056)
									-0.992	(0.051)
R015808a**	RI	8	0.843	(0.068)	-0.938	(0.117)	0.218	(0.040)		
R015809a**	RI	9	0.639	(0.026)	0.247	(0.042)	0.000	(0.000)	1.334	(0.064)
									-1.334	(0.063)

***Note:** a = item parameters are based on only 1994 data; b = item parameters are based on only 1992 data.*

***Note:** This block name identifies two different sets of items for 1992 and 1994–a=1994 items; b=1992 items.*

Table D-2

IRT Parameters for the Reading Main Samples
Reading to Gain Information Scale, Age 9/Grade 4

NAEP ID	Block	Item	A	S.E.	B	S.E.	C	S.E.	D	S.E.
R012201a*	RF	1	0.272	(0.024)	-0.484	(0.100)	0.000	(0.000)		
R012201b*	RF	1	0.399	(0.029)	0.142	(0.064)	0.000	(0.000)		
R012202	RF	2	0.819	(0.060)	0.534	(0.059)	0.224	(0.021)		
R012203	RF	3	0.797	(0.061)	0.682	(0.056)	0.200	(0.020)		
R012204	RF	4	0.523	(0.015)	0.385	(0.021)	0.000	(0.000)	1.114	(0.046)
									-0.256	(0.045)
									-0.858	(0.057)
R012205	RF	5	1.331	(0.090)	0.647	(0.036)	0.264	(0.015)		
R012206	RF	6	1.309	(0.046)	0.800	(0.021)	0.000	(0.000)		
R012207	RF	7	0.573	(0.043)	-0.530	(0.137)	0.222	(0.041)		
R012208	RF	8	0.906	(0.033)	-0.197	(0.027)	0.000	(0.000)		
R012209a*	RF	9	1.412	(0.106)	0.475	(0.041)	0.146	(0.018)		
R012209b*	RF	9	1.724	(0.162)	0.515	(0.044)	0.264	(0.021)		
R012210	RF	10	0.767	(0.032)	-0.965	(0.047)	0.000	(0.000)		
R012301b**	RH	1	0.805	(0.079)	-0.020	(0.112)	0.293	(0.039)		
R012302b**	RH	2	1.052	(0.074)	-0.325	(0.069)	0.204	(0.032)		
R012303b**	RH	3	1.170	(0.051)	-0.260	(0.028)	0.000	(0.000)		
R012304b**	RH	4	1.716	(0.323)	2.188	(0.200)	0.244	(0.010)		
R012305b**	RH	5	0.527	(0.018)	1.131	(0.036)	0.000	(0.000)	2.450	(0.066)
									0.130	(0.058)
									-2.580	(0.176)
R012306b**	RH	6	0.834	(0.047)	0.940	(0.046)	0.000	(0.000)		
R012307b**	RH	7	1.384	(0.093)	-0.001	(0.046)	0.172	(0.024)		
R012308b**	RH	8	0.913	(0.048)	0.540	(0.035)	0.000	(0.000)		
R012309b**	RH	9	0.920	(0.123)	0.985	(0.078)	0.249	(0.027)		
R012310b**	RH	10	0.916	(0.054)	0.620	(0.039)	0.000	(0.000)		
R012501	RJ	1	0.539	(0.172)	3.749	(0.773)	0.285	(0.014)		
R012502	RJ	2	1.153	(0.061)	-1.382	(0.073)	0.218	(0.034)		
R012503a*	RJ	3	1.219	(0.053)	0.177	(0.025)	0.000	(0.000)		
R012503b*	RJ	3	1.114	(0.052)	0.626	(0.030)	0.000	(0.000)		
R012504	RJ	4	0.861	(0.029)	0.042	(0.023)	0.000	(0.000)		
R012505	RJ	5	1.274	(0.063)	-0.559	(0.047)	0.222	(0.023)		
R012506	RJ	6	0.928	(0.031)	0.077	(0.022)	0.000	(0.000)		
R012507	RJ	7	1.322	(0.072)	-0.304	(0.047)	0.294	(0.022)		
R012508	RJ	8	1.125	(0.037)	-0.135	(0.021)	0.000	(0.000)		
R012509	RJ	9	0.631	(0.045)	-0.727	(0.134)	0.240	(0.042)		
R012510	RJ	10	0.962	(0.062)	-0.264	(0.072)	0.303	(0.028)		
R012511	RJ	11	1.129	(0.040)	-0.326	(0.024)	0.000	(0.000)		
R012512a*	RJ	12	0.409	(0.021)	0.687	(0.039)	0.000	(0.000)	0.910	(0.088)
									0.195	(0.088)
									-1.106	(0.115)

***Note**: *a = item parameters are based on only 1994 data; b = item parameters are based on only 1992 data.*
****Note**: *This block name identifies two different sets of items for 1992 and 1994–a=1994 items; b=1992 items.*

Table D-2 (continued)

IRT Parameters for the Reading Main Samples
Reading to Gain Information Scale, Age 9/Grade 4

NAEP ID	Block	Item	A	S.E.	B	S.E.	C	S.E.	D	S.E.
R012512b*	RJ	12	0.451	(0.024)	1.140	(0.044)	0.000	(0.000)	0.842	(0.076)
									0.390	(0.085)
									-1.232	(0.136)
R012701	RG	1	1.351	(0.072)	0.107	(0.037)	0.262	(0.018)		
R012702	RG	2	0.635	(0.024)	-0.950	(0.044)	0.000	(0.000)		
R012703	RG	3	1.168	(0.040)	0.791	(0.022)	0.000	(0.000)		
R012704	RG	4	1.716	(0.099)	0.913	(0.024)	0.169	(0.010)		
R012705a*	RG	5	1.608	(0.094)	1.330	(0.036)	0.000	(0.000)		
R012705b*	RG	5	1.283	(0.102)	1.820	(0.080)	0.000	(0.000)		
R012706a*	RG	6	0.648	(0.046)	1.543	(0.089)	0.000	(0.000)		
R012706b*	RG	6	0.828	(0.048)	0.957	(0.047)	0.000	(0.000)		
R012707	RG	7	2.473	(0.147)	0.559	(0.022)	0.267	(0.012)		
R012708a*	RG	8	0.742	(0.035)	1.852	(0.037)	0.000	(0.000)	1.250	(0.044)
									0.382	(0.063)
									-1.633	(0.231)
R012708b*	RG	8	1.050	(0.048)	1.889	(0.031)	0.000	(0.000)	1.343	(0.032)
									-0.118	(0.065)
									-1.226	(0.297)
R012709	RG	9	0.591	(0.053)	0.034	(0.127)	0.229	(0.039)		
R012710a*	RG	10	1.064	(0.068)	1.004	(0.044)	0.000	(0.000)		
R012710b*	RG	10	1.141	(0.069)	0.779	(0.038)	0.000	(0.000)		
R015701a**	RH	1	1.000	(0.081)	-0.691	(0.102)	0.299	(0.039)		
R015702a**	RH	2	0.697	(0.022)	0.229	(0.033)	0.000	(0.000)	1.405	(0.049)
									-1.405	(0.051)
R015703a**	RH	3	0.763	(0.025)	0.216	(0.031)	0.000	(0.000)	1.325	(0.046)
									-1.325	(0.047)
R015704a**	RH	4	0.728	(0.031)	0.108	(0.026)	0.000	(0.000)	0.326	(0.047)
									-0.326	(0.044)
R015705a**	RH	5	0.851	(0.036)	0.395	(0.026)	0.000	(0.000)	0.664	(0.040)
									-0.664	(0.042)
R015706a**	RH	6	1.337	(0.140)	1.242	(0.053)	0.230	(0.015)		
R015707a**	RH	7	0.619	(0.025)	0.519	(0.036)	0.000	(0.000)	1.112	(0.053)
									-1.112	(0.061)
R015708a**	RH	8	0.706	(0.065)	0.123	(0.096)	0.171	(0.033)		
R015709a**	RH	9	0.583	(0.036)	1.215	(0.054)	0.000	(0.000)	0.324	(0.060)
									-0.324	(0.085)

*_Note: a = item parameters are based on only 1994 data; b = item parameters are based on only 1992 data._

**_Note: This block name identifies two different sets of items for 1992 and 1994–a=1994 items; b=1992 items._

BEST COPY AVAILABLE

Table D-3
IRT Parameters for the Reading Main Samples
Reading for Literary Experience Scale, Age 13/Grade 8

NAEP ID	Block	Item	A	S.E.	B	S.E.	C	S.E.	D	S.E.
R012601	RE	1	0.888	(0.028)	0.200	(0.023)	0.000	(0.000)		
R012602	RE	2	0.948	(0.046)	0.335	(0.038)	0.099	(0.015)		
R012603	RE	3	1.357	(0.066)	-0.798	(0.050)	0.208	(0.028)		
R012604	RE	4	0.799	(0.026)	0.260	(0.025)	0.000	(0.000)		
R012605	RE	5	0.703	(0.041)	-0.230	(0.086)	0.174	(0.032)		
R012606	RE	6	1.514	(0.077)	-0.593	(0.044)	0.250	(0.025)		
R012607a*	RE	7	0.668	(0.025)	0.585	(0.024)	0.000	(0.000)	0.857	(0.050)
									0.096	(0.051)
									-0.952	(0.066)
R012607b*	RE	7	0.722	(0.024)	1.171	(0.025)	0.000	(0.000)	1.384	(0.046)
									0.830	(0.044)
									-2.213	(0.128)
R012608	RE	8	0.599	(0.037)	-1.518	(0.164)	0.236	(0.057)		
R012609	RE	9	1.213	(0.064)	-0.161	(0.047)	0.226	(0.022)		
R012610	RE	10	1.268	(0.078)	-0.264	(0.059)	0.375	(0.025)		
R012611a*	RE	11	0.787	(0.039)	-0.817	(0.048)	0.000	(0.000)		
R012611b*	RE	11	0.715	(0.041)	-1.492	(0.078)	0.000	(0.000)		
R012801b**	RD	1	0.607	(0.049)	-0.151	(0.116)	0.166	(0.037)		
R012802b**	RD	2	1.295	(0.085)	-1.054	(0.071)	0.203	(0.038)		
R012803b**	RD	3	0.864	(0.040)	0.632	(0.037)	0.000	(0.000)		
R012804b**	RD	4	0.813	(0.073)	-0.363	(0.122)	0.282	(0.044)		
R012805b**	RD	5	0.823	(0.069)	0.196	(0.081)	0.190	(0.030)		
R012806b**	RD	6	0.972	(0.069)	-0.534	(0.083)	0.210	(0.036)		
R012807b**	RD	7	0.669	(0.036)	0.894	(0.053)	0.000	(0.000)		
R012808b**	RD	8	0.743	(0.033)	1.777	(0.043)	0.000	(0.000)	0.401	(0.048)
									0.338	(0.074)
									-0.739	(0.139)
R012809b**	RD	9	0.893	(0.067)	-0.695	(0.104)	0.240	(0.042)		
R012810b**	RD	10	0.741	(0.035)	-0.087	(0.038)	0.000	(0.000)		
R012811b**	RD	11	1.019	(0.077)	-0.608	(0.091)	0.240	(0.040)		
R012812b**	RD	12	0.814	(0.082)	0.086	(0.109)	0.301	(0.037)		
R012813b**	RD	13	0.584	(0.036)	0.643	(0.057)	0.000	(0.000)		
R013101	RC	1	0.818	(0.046)	0.432	(0.047)	0.113	(0.018)		
R013102a*	RC	2	0.657	(0.032)	0.031	(0.041)	0.000	(0.000)		
R013102b*	RC	2	0.761	(0.039)	0.905	(0.047)	0.000	(0.000)		
R013103	RC	3	0.688	(0.047)	-0.558	(0.125)	0.260	(0.043)		
R013104	RC	4	0.957	(0.029)	-0.260	(0.022)	0.000	(0.000)		
R013105	RC	5	0.829	(0.031)	1.116	(0.035)	0.000	(0.000)		
R013106	RC	6	1.000	(0.024)	1.665	(0.018)	0.000	(0.000)	2.150	(0.023)
									0.106	(0.030)
									-2.256	(0.211)

**__Note__: a = item parameters are based on only 1994 data; b = item parameters are based on only 1992 data.*
***__Note__: This block name identifies two different sets of items for 1992 and 1994–a=1994 items; b=1992 items.*

Table D-3 (continued)

IRT Parameters for the Reading Main Samples
Reading for Literary Experience Scale, Age 13/Grade 8

NAEP ID	Block	Item	A	S.E.	B	S.E.	C	S.E.	D	S.E.
R013107	RC	7	1.159	(0.076)	0.418	(0.045)	0.289	(0.018)		
R013108	RC	8	0.683	(0.095)	3.616	(0.399)	0.000	(0.000)		
R013109	RC	9	1.204	(0.077)	0.956	(0.033)	0.144	(0.011)		
R013110	RC	10	0.476	(0.023)	0.533	(0.046)	0.000	(0.000)		
R013111	RC	11	0.713	(0.039)	1.882	(0.077)	0.000	(0.000)		
R015901a**	RD	1	0.586	(0.031)	0.117	(0.045)	0.000	(0.000)		
R015902a**	RD	2	0.702	(0.027)	0.411	(0.030)	0.000	(0.000)	0.938	(0.043)
									-0.938	(0.051)
R015903a**	RD	3	0.873	(0.079)	0.404	(0.076)	0.214	(0.028)		
R015904a**	RD	4	0.660	(0.029)	1.688	(0.045)	0.000	(0.000)	1.148	(0.042)
									-1.148	(0.101)
R015905a**	RD	5	0.552	(0.025)	0.616	(0.037)	0.000	(0.000)	0.522	(0.054)
									-0.522	(0.065)
R015906a**	RD	6	0.551	(0.021)	2.099	(0.040)	0.000	(0.000)	2.482	(0.052)
									0.085	(0.074)
									-2.568	(0.351)
R015907a**	RD	7	0.503	(0.018)	0.674	(0.049)	0.000	(0.000)	1.797	(0.066)
									-1.797	(0.088)
R015908a**	RD	8	0.767	(0.040)	1.136	(0.041)	0.000	(0.000)	0.749	(0.047)
									-0.749	(0.074)

**Note: a = item parameters are based on only 1994 data; b = item parameters are based on only 1992 data.*

***Note: This block name identifies two different sets of items for 1992 and 1994–a=1994 items; b=1992 items.*

Table D-4

IRT Parameters for the Reading Main Samples
Reading to Gain Information Scale, Age 13/Grade 8

NAEP ID	Block	Item	A	S.E.	B	S.E.	C	S.E.	D	S.E.
R012701	RG	1	1.187	(0.068)	-1.091	(0.076)	0.349	(0.033)		
R012702	RG	2	0.656	(0.026)	-1.893	(0.065)	0.000	(0.000)		
R012703	RG	8	0.967	(0.031)	-0.098	(0.022)	0.000	(0.000)		
R012704	RG	4	1.363	(0.066)	-0.290	(0.039)	0.228	(0.019)		
R012705a*	RG	5	1.110	(0.049)	0.200	(0.026)	0.000	(0.000)		
R012705b*	RG	5	1.036	(0.049)	0.558	(0.029)	0.000	(0.000)		
R012706a*	RG	6	0.536	(0.034)	0.633	(0.056)	0.000	(0.000)		
R012706b*	RG	6	0.509	(0.031)	0.161	(0.050)	0.000	(0.000)		
R012707	RG	3	2.137	(0.119)	-0.519	(0.034)	0.361	(0.020)		
R012708a*	RG	10	0.583	(0.022)	0.788	(0.030)	0.000	(0.000)	1.544	(0.057)
									0.124	(0.054)
									-1.668	(0.104)
R012708b*	RG	10	0.652	(0.021)	1.236	(0.032)	0.000	(0.000)	2.005	(0.048)
									-0.296	(0.054)
									-1.709	(0.154)
R012709	RG	9	0.816	(0.058)	-0.523	(0.105)	0.367	(0.034)		
R012710a*	RG	11	0.880	(0.042)	-0.378	(0.038)	0.000	(0.000)		
R012710b*	RG	11	1.059	(0.047)	-0.622	(0.037)	0.000	(0.000)		
R012711	RG	7	1.349	(0.064)	-0.152	(0.035)	0.194	(0.018)		
R012712	RG	12	1.098	(0.082)	0.379	(0.051)	0.345	(0.020)		
R012713	RG	13	1.311	(0.044)	-0.407	(0.022)	0.000	(0.000)		
R012901b**	RH	1	0.815	(0.039)	0.096	(0.034)	0.000	(0.000)		
R012902b**	RH	2	0.457	(0.049)	-0.458	(0.215)	0.256	(0.051)		
R012903b**	RH	3	0.645	(0.026)	0.898	(0.028)	0.000	(0.000)	1.066	(0.047)
									-0.129	(0.054)
									-0.937	(0.087)
R012904b**	RH	4	1.309	(0.141)	0.883	(0.048)	0.272	(0.018)		
R012905b**	RH	5	0.810	(0.040)	0.405	(0.035)	0.000	(0.000)		
R012906b**	RH	6	1.596	(0.102)	0.100	(0.037)	0.177	(0.019)		
R012907b**	RH	7	0.732	(0.038)	0.217	(0.037)	0.000	(0.000)		
R012908b**	RH	8	1.388	(0.106)	-0.121	(0.058)	0.296	(0.027)		
R012909b**	RH	9	0.997	(0.045)	-0.629	(0.038)	0.000	(0.000)		
R012910b**	RH	10	0.836	(0.067)	1.869	(0.108)	0.000	(0.000)		
R012911b**	RH	11	1.574	(0.116)	0.499	(0.034)	0.164	(0.016)		
R012912b**	RH	12	1.441	(0.111)	-0.668	(0.070)	0.310	(0.033)		
R012913b**	RH	13	1.292	(0.108)	-0.316	(0.076)	0.344	(0.032)		
R012914b**	RH	14	0.738	(0.045)	0.542	(0.043)	0.000	(0.000)		
R013201	RF	1	0.711	(0.020)	0.981	(0.019)	0.000	(0.000)	0.990	(0.030)
									-0.022	(0.035)
									-0.967	(0.060)
R013202	RF	2	0.714	(0.045)	-0.255	(0.085)	0.199	(0.030)		

**Note:* *a = item parameters are based on only 1994 data; b = item parameters are based on only 1992 data.*

***Note: This block name identifies two different sets of items for 1992 and 1994–a=1994 items; b=1992 items.*

Table D-4 (continued)

IRT Parameters for the Reading Main Samples
Reading to Gain Information Scale, Age 13/Grade 8

NAEP ID	Block	Item	A	S.E.	B	S.E.	C	S.E.	D	S.E.
R013203	RF	3	1.147	(0.040)	-1.451	(0.037)	0.000	(0.000)		
R013204	RF	4	1.279	(0.079)	0.088	(0.046)	0.352	(0.019)		
R013205a*	RF	5	1.124	(0.051)	-0.941	(0.041)	0.000	(0.000)		
R013205b*	RF	5	0.688	(0.035)	-0.353	(0.044)	0.000	(0.000)		
R013206	RF	6	0.866	(0.052)	0.213	(0.051)	0.160	(0.021)		
R013207	RF	7	0.791	(0.027)	-0.331	(0.028)	0.000	(0.000)		
R013208	RF	8	1.998	(0.100)	0.060	(0.025)	0.238	(0.015)		
R013209	RF	9	0.759	(0.035)	1.371	(0.049)	0.000	(0.000)		
R013210	RF	10	1.500	(0.130)	1.735	(0.068)	0.263	(0.008)		
R013211a*	RF	11	0.547	(0.039)	0.969	(0.068)	0.000	(0.000)		
R013211b*	RF	11	0.759	(0.042)	0.245	(0.038)	0.000	(0.000)		
R013212	RF	12	0.595	(0.018)	2.232	(0.033)	0.000	(0.000)	1.979	(0.035)
									-0.581	(0.075)
									-1.398	(0.269)
R014701b**	RM	1	0.433	(0.102)	1.540	(0.241)	0.387	(0.043)		
R014702b**	RM	2	0.844	(0.039)	-0.200	(0.035)	0.000	(0.000)		
R014703b**	RM	3	0.909	(0.057)	1.398	(0.062)	0.000	(0.000)		
R014704b**	RM	4	1.087	(0.047)	0.124	(0.027)	0.000	(0.000)		
R014705b**	RM	5	0.547	(0.023)	0.554	(0.026)	0.000	(0.000)	0.552	(0.060)
									0.163	(0.063)
									-0.715	(0.075)
R014706b**	RM	6	1.036	(0.045)	-0.372	(0.032)	0.000	(0.000)		
R014707b**	RM	7	1.179	(0.117)	0.909	(0.047)	0.188	(0.018)		
R014708b**	RM	8	0.665	(0.058)	-0.626	(0.144)	0.250	(0.045)		
R014709b**	RM	9	0.975	(0.043)	-0.509	(0.036)	0.000	(0.000)		
R014710b**	RM	10	0.729	(0.045)	1.162	(0.060)	0.000	(0.000)		
R014711b**	RM	11	0.873	(0.086)	-0.176	(0.114)	0.352	(0.038)		
R014712b**	RM	12	1.011	(0.214)	1.823	(0.163)	0.296	(0.017)		
R014713b**	RM	13	0.663	(0.024)	0.880	(0.029)	0.000	(0.000)	1.351	(0.046)
									-0.682	(0.059)
									-0.668	(0.091)
R016001a**	RH	1	0.282	(0.025)	-0.835	(0.114)	0.000	(0.000)		
R016002a**	RH	2	0.792	(0.024)	0.562	(0.024)	0.000	(0.000)	1.596	(0.044)
									-0.769	(0.045)
									-0.827	(0.069)
R016003a**	RH	3	0.574	(0.025)	0.066	(0.033)	0.000	(0.000)	0.829	(0.056)
									-0.829	(0.053)
R016004a**	RH	4	0.503	(0.021)	0.773	(0.042)	0.000	(0.000)	1.253	(0.057)
									-1.253	(0.076)
R016005a**	RH	5	0.850	(0.066)	-0.271	(0.087)	0.202	(0.033)		

__Note__: a = item parameters are based on only 1994 data; b = item parameters are based on only 1992 data.

***__Note__: This block name identifies two different sets of items for 1992 and 1994–a=1994 items; b=1992 items.*

Table D-4 (continued)

IRT Parameters for the Reading Main Samples
Reading to Gain Information Scale, Age 13/Grade 8

NAEP ID	Block	Item	A	S.E.	B	S.E.	C	S.E.	D	S.E.
R016006a**	RH	6	0.435	(0.021)	-0.093	(0.039)	0.000	(0.000)	-0.017	(0.076)
									0.017	(0.071)
R016007a**	RH	7	1.029	(0.101)	0.441	(0.067)	0.266	(0.027)		
R016008a**	RH	8	1.443	(0.124)	-0.394	(0.075)	0.403	(0.032)		
R016009a**	RH	9	0.486	(0.023)	0.926	(0.037)	0.000	(0.000)	0.564	(0.064)
									-0.302	(0.082)
									-0.262	(0.104)
R016201a**	RM	1	0.540	(0.029)	-3.053	(0.127)	0.000	(0.000)	0.006	(0.193)
									-0.006	(0.089)
R016202a**	RM	2	0.743	(0.022)	-0.585	(0.040)	0.000	(0.000)	2.180	(0.079)
									-2.180	(0.046)
R016203a**	RM	3	0.485	(0.055)	-0.448	(0.220)	0.292	(0.052)		
R016204a**	RM	4	0.618	(0.017)	-0.326	(0.038)	0.000	(0.000)	2.922	(0.124)
									0.483	(0.047)
									-3.405	(0.097)
R016205a**	RM	5	0.688	(0.022)	0.812	(0.034)	0.000	(0.000)	1.482	(0.043)
									-1.482	(0.068)
R016206a**	RM	6	0.970	(0.074)	-0.674	(0.097)	0.294	(0.038)		
R016207a**	RM	7	0.618	(0.028)	0.400	(0.029)	0.000	(0.000)	0.240	(0.050)
									-0.240	(0.054)
R016208a**	RM	8	1.064	(0.084)	-0.612	(0.093)	0.319	(0.037)		
R016209a**	RM	9	1.285	(0.084)	-0.526	(0.060)	0.218	(0.029)		
R016210a**	RM	10	0.723	(0.026)	1.063	(0.026)	0.000	(0.000)	1.491	(0.045)
									0.241	(0.045)
									-1.732	(0.116)
R016211a**	RM	11	0.541	(0.021)	-1.650	(0.045)	0.000	(0.000)	2.038	(0.130)
									-2.038	(0.049)
R016212a**	RM	12	0.368	(0.021)	0.200	(0.044)	0.000	(0.000)	0.166	(0.084)
									-0.166	(0.085)
R016213a**	RM	13	0.439	(0.025)	-1.645	(0.078)	0.000	(0.000)	0.831	(0.120)
									-0.831	(0.065)

Note: *a = item parameters are based on only 1994 data; b = item parameters are based on only 1992 data.*

Note: *This block name identifies two different sets of items for 1992 and 1994–a=1994 items; b=1992 items.*

Table D-5

IRT Parameters for the Reading Main Samples
Reading to Perform a Task Scale, Age 13/Grade 8

NAEP ID	Block	Item	A	S.E.	B	S.E.	C	S.E.	D	S.E.
R013001	RK	1	1.070	(0.034)	-0.713	(0.026)	0.000	(0.000)		
R013002	RK	2	1.580	(0.070)	-0.129	(0.028)	0.164	(0.016)		
R013003	RK	3	1.007	(0.032)	-0.265	(0.022)	0.000	(0.000)		
R013004a*	RK	4	0.424	(0.020)	0.798	(0.045)	0.000	(0.000)	-0.325	(0.072)
									0.325	(0.084)
R013004b*	RK	4	0.475	(0.026)	1.049	(0.052)	0.000	(0.000)	0.257	(0.061)
									-0.257	(0.082)
R013005	RK	5	0.839	(0.029)	-0.754	(0.033)	0.000	(0.000)		
R013006	RK	6	0.994	(0.058)	-0.086	(0.056)	0.249	(0.024)		
R013007	RK	7	0.745	(0.028)	-0.997	(0.041)	0.000	(0.000)		
R013008	RK	8	0.855	(0.030)	0.135	(0.024)	0.000	(0.000)		
R013009a*	RK	9	1.116	(0.055)	-0.926	(0.044)	0.000	(0.000)		
R013009b*	RK	9	0.859	(0.043)	-0.523	(0.042)	0.000	(0.000)		
R013010	RK	10	0.857	(0.031)	-0.738	(0.034)	0.000	(0.000)		
R013011a*	RK	11	0.599	(0.036)	-0.195	(0.050)	0.000	(0.000)		
R013011b*	RK	11	0.617	(0.039)	0.536	(0.051)	0.000	(0.000)		
R013012	RK	12	1.351	(0.093)	0.320	(0.043)	0.319	(0.019)		
R013301b**	RI	1	0.650	(0.079)	0.028	(0.163)	0.382	(0.043)		
R013302b**	RI	2	1.085	(0.071)	1.633	(0.068)	0.000	(0.000)		
R013303b**	RI	3	0.950	(0.092)	0.361	(0.076)	0.285	(0.029)		
R013304b**	RI	4	0.518	(0.046)	2.286	(0.175)	0.000	(0.000)		
R013305b**	RI	5	1.425	(0.123)	-0.995	(0.090)	0.428	(0.040)		
R013306b**	RI	6	0.689	(0.067)	-0.137	(0.128)	0.278	(0.041)		
R013307b**	RI	7	1.007	(0.051)	0.854	(0.036)	0.000	(0.000)		
R013308b**	RI	8	1.432	(0.150)	0.598	(0.056)	0.408	(0.021)		
R013309b**	RI	9	1.076	(0.142)	1.251	(0.068)	0.251	(0.019)		
R013310b**	RI	10	0.809	(0.044)	0.838	(0.043)	0.000	(0.000)		
R013311b**	RI	11	1.099	(0.110)	0.338	(0.074)	0.352	(0.028)		
R013312b**	RI	12	0.540	(0.030)	2.029	(0.078)	0.000	(0.000)	-0.053	(0.069)
									0.397	(0.113)
									-0.344	(0.198)
R013401	RJ	1	1.092	(0.062)	0.415	(0.036)	0.176	(0.016)		
R013402a*	RJ	2	0.929	(0.043)	0.345	(0.031)	0.000	(0.000)		
R013402b*	RJ	2	0.833	(0.040)	0.102	(0.033)	0.000	(0.000)		
R013403	RJ	3	0.565	(0.012)	0.669	(0.021)	0.000	(0.000)	-1.786	(0.064)
									1.786	(0.067)
R013404	RJ	4	1.165	(0.080)	0.221	(0.052)	0.372	(0.020)		
R013405	RJ	5	1.030	(0.032)	-0.051	(0.020)	0.000	(0.000)		
R013406	RJ	6	0.759	(0.030)	0.790	(0.032)	0.000	(0.000)		
R013407	RJ	7	0.713	(0.026)	-0.392	(0.031)	0.000	(0.000)		
R013408	RJ	8	0.796	(0.060)	0.357	(0.067)	0.254	(0.024)		

__Note__: a = item parameters are based on only 1994 data; b = item parameters are based on only 1992 data.
**__Note__: This block name identifies two different sets of items for 1992 and 1994–a=1994 items; b=1992 items.*

BEST COPY AVAILABLE

Table D-5 (continued)

IRT Parameters for the Reading Main Samples
Reading to Perform a Task Scale, Age 13/Grade 8

NAEP ID	Block	Item	A	S.E.	B	S.E.	C	S.E.	D	S.E.
R013409	RJ	9	0.790	(0.028)	-0.136	(0.026)	0.000	(0.000)		
R013410	RJ	10	0.832	(0.057)	-0.346	(0.089)	0.298	(0.032)		
R013411a*	RJ	11	0.658	(0.038)	-0.519	(0.053)	0.000	(0.000)		
R013411b*	RJ	11	0.730	(0.046)	0.962	(0.056)	0.000	(0.000)		
R013412	RJ	12	0.435	(0.025)	-1.585	(0.096)	0.000	(0.000)		
R016101	RI	1	0.615	(0.026)	0.156	(0.028)	0.000	(0.000)	0.161	(0.052)
									-0.161	(0.052)
R016102a**	RI	2	1.057	(0.085)	-0.370	(0.086)	0.323	(0.034)		
R016103a**	RI	3	1.658	(0.153)	0.618	(0.041)	0.307	(0.018)		
R016104a**	RI	4	0.825	(0.042)	-1.348	(0.062)	0.000	(0.000)		
R016105a**	RI	5	1.534	(0.102)	-0.371	(0.050)	0.245	(0.027)		
R016106a**	RI	6	1.874	(0.135)	1.018	(0.034)	0.217	(0.012)		
R016107a**	RI	7	0.851	(0.029)	0.107	(0.022)	0.000	(0.000)	-0.154	(0.043)
									0.154	(0.042)
R016108a**	RI	8	0.405	(0.017)	0.336	(0.039)	0.000	(0.000)	-0.843	(0.085)
									0.843	(0.088)
R016109a**	RI	9	0.497	(0.018)	0.870	(0.045)	0.000	(0.000)	1.691	(0.059)
									-1.691	(0.087)

******Note*:* *a = item parameters are based on only 1994 data; b = item parameters are based on only 1992 data.*

*******Note*: *This block name identifies two different sets of items for 1992 and 1994–a=1994 items; b=1992 items.*

Table D-6

IRT Parameters for the Reading Main Samples
Reading for Literary Experience Scale, Age 17/Grade 12

NAEP ID	Block	Item	A	S.E.	B	S.E.	C	S.E.	D	S.E.
R013101	RC	1	0.786	(0.042)	-0.239	(0.062)	0.118	(0.024)		
R013102	RC	2	0.971	(0.046)	0.196	(0.032)	0.000	(0.000)		
R013103	RC	3	0.687	(0.045)	-1.190	(0.140)	0.247	(0.050)		
R013104	RC	4	1.244	(0.041)	-0.816	(0.025)	0.000	(0.000)		
R013105	RC	5	0.845	(0.029)	-0.190	(0.026)	0.000	(0.000)		
R013106	RC	6	1.048	(0.026)	0.766	(0.016)	0.000	(0.000)	1.732	(0.028)
									0.195	(0.024)
									-1.927	(0.071)
R013107	RC	7	1.056	(0.064)	-0.422	(0.067)	0.270	(0.029)		
R013108	RC	8	0.683	(0.078)	2.819	(0.259)	0.000	(0.000)		
R013109	RC	9	1.171	(0.061)	0.246	(0.034)	0.112	(0.015)		
R013110	RC	10	0.504	(0.025)	0.018	(0.042)	0.000	(0.000)		
R013111	RC	11	0.742	(0.034)	0.977	(0.042)	0.000	(0.000)		
R013501	RD	1	0.875	(0.030)	-0.341	(0.027)	0.000	(0.000)		
R013502	RD	2	1.252	(0.069)	-0.604	(0.056)	0.242	(0.028)		
R013503	RD	3	0.520	(0.035)	-1.052	(0.085)	0.000	(0.000)		
R013504	RD	4	0.619	(0.040)	-0.287	(0.098)	0.159	(0.033)		
R013505	RD	5	0.811	(0.046)	-1.102	(0.062)	0.000	(0.000)		
R013506	RD	6	0.387	(0.012)	1.694	(0.057)	0.000	(0.000)	-2.274	(0.106)
									2.274	(0.123)
R013507	RD	7	0.895	(0.059)	0.223	(0.057)	0.166	(0.022)		
R013508	RD	8	0.421	(0.040)	0.804	(0.099)	0.000	(0.000)		
R013509	RD	9	0.827	(0.055)	0.174	(0.048)	0.000	(0.000)		
R013601b**	RE	1	1.258	(0.108)	0.020	(0.067)	0.254	(0.030)		
R013602b**	RE	2	0.705	(0.038)	0.238	(0.041)	0.000	(0.000)		
R013603b**	RE	3	1.158	(0.090)	-0.134	(0.069)	0.211	(0.032)		
R013604b**	RE	4	0.499	(0.035)	-1.281	(0.096)	0.000	(0.000)		
R013605b**	RE	5	0.812	(0.041)	0.041	(0.037)	0.000	(0.000)		
R013606b**	RE	6	0.591	(0.079)	0.661	(0.137)	0.230	(0.041)		
R013607b**	RE	7	0.689	(0.038)	-0.624	(0.052)	0.000	(0.000)		
R013608b**	RE	8	0.938	(0.110)	0.677	(0.079)	0.274	(0.028)		
R013609b**	RE	9	0.933	(0.048)	-0.799	(0.046)	0.000	(0.000)		
R013610b**	RE	10	0.957	(0.053)	0.751	(0.039)	0.000	(0.000)		
R013611b**	RE	11	0.698	(0.043)	0.692	(0.051)	0.000	(0.000)		
R013612b**	RE	12	0.876	(0.095)	0.455	(0.085)	0.209	(0.032)		
R016301a**	RE	1	0.431	(0.016)	0.140	(0.052)	0.000	(0.000)	1.856	(0.079)
									-1.856	(0.083)
R016302a**	RE	2	0.461	(0.020)	0.410	(0.046)	0.000	(0.000)	1.279	(0.068)
									-1.279	(0.079)

__Note__: a = item parameters are based on only 1994 data; b = item parameters are based on only 1992 data.

**__Note__: *This block name identifies two different sets of items for 1992 and 1994–a=1994 items; b=1992 items.*

Table D-6 (continued)

IRT Parameters for the Reading Main Samples
Reading for Literary Experience Scale, Age 17/Grade 12

NAEP ID	Block	Item	A	S.E.	B	S.E.	C	S.E.	D	S.E.
R016303a**	RE	3	0.588	(0.027)	0.856	(0.041)	0.000	(0.000)	0.937	(0.053)
									-0.937	(0.073)
R016304a**	RE	4	1.330	(0.120)	0.123	(0.063)	0.337	(0.026)		
R016305a**	RE	5	0.531	(0.027)	-0.323	(0.041)	0.000	(0.000)	0.612	(0.072)
									-0.612	(0.061)
R016306a**	RE	6	0.571	(0.029)	0.580	(0.039)	0.000	(0.000)	0.372	(0.060)
									-0.372	(0.070)
R016307a**	RE	7	0.609	(0.033)	1.053	(0.047)	0.000	(0.000)	0.794	(0.057)
									-0.794	(0.084)
R016308a**	RE	8	0.452	(0.027)	1.744	(0.076)	0.000	(0.000)	0.405	(0.081)
									-0.318	(0.134)
									-0.087	(0.195)

****Note:*** *a = item parameters are based on only 1994 data; b = item parameters are based on only 1992 data.*

****Note*:** *This block name identifies two different sets of items for 1992 and 1994–a=1994 items; b=1992 items.*

Table D-7

IRT Parameters for the Reading Main Samples
Reading to Gain Information Scale, Age 17/Grade 12

NAEP ID	Block	Item	A	S.E.	B	S.E.	C	S.E.	D	S.E.
R013201	RF	1	0.548	(0.023)	0.385	(0.030)	0.000	(0.000)	1.269	(0.069)
									0.011	(0.059)
									-1.279	(0.082)
R013202	RF	2	0.696	(0.051)	-0.922	(0.139)	0.310	(0.044)		
R013203	RF	3	0.944	(0.042)	-2.260	(0.069)	0.000	(0.000)		
R013204	RF	4	0.827	(0.055)	-1.171	(0.120)	0.323	(0.042)		
R013205	RF	5	0.501	(0.035)	-1.383	(0.105)	0.000	(0.000)		
R013206	RF	6	0.653	(0.044)	-0.539	(0.108)	0.204	(0.036)		
R013207	RF	7	0.659	(0.028)	-1.317	(0.056)	0.000	(0.000)		
R013208	RF	8	1.693	(0.098)	-0.904	(0.049)	0.293	(0.028)		
R013209	RF	9	0.591	(0.040)	0.546	(0.052)	0.000	(0.000)		
R013210	RF	10	0.843	(0.104)	1.448	(0.076)	0.227	(0.018)		
R013211	RF	11	0.711	(0.044)	-0.881	(0.067)	0.000	(0.000)		
R013212	RF	12	0.470	(0.018)	2.017	(0.053)	0.000	(0.000)	2.406	(0.065)
									-1.013	(0.118)
									-1.393	(0.324)
R013701	RG	1	0.962	(0.055)	-1.425	(0.068)	0.000	(0.000)		
R013702	RG	2	1.167	(0.057)	-0.699	(0.039)	0.000	(0.000)		
R013703	RG	3	0.955	(0.064)	-1.556	(0.122)	0.333	(0.048)		
R013704	RG	4	0.847	(0.053)	1.036	(0.053)	0.000	(0.000)		
R013705	RG	5	1.131	(0.065)	-1.110	(0.076)	0.273	(0.035)		
R013706	RG	6	0.524	(0.036)	-0.546	(0.066)	0.000	(0.000)		
R013707	RG	7	0.750	(0.043)	0.033	(0.057)	0.114	(0.022)		
R013708	RG	8	0.498	(0.036)	-1.179	(0.100)	0.000	(0.000)		
R013709	RG	9	0.427	(0.056)	0.960	(0.152)	0.211	(0.038)		
R013710	RG	10	0.508	(0.039)	0.595	(0.064)	0.000	(0.000)		
R013711	RG	11	0.433	(0.075)	1.490	(0.167)	0.258	(0.038)		
R013712	RG	12	0.906	(0.056)	-0.610	(0.053)	0.000	(0.000)		
R013801b**	RH	1	0.948	(0.048)	-0.254	(0.036)	0.000	(0.000)		
R013802b**	RH	2	0.861	(0.117)	0.951	(0.085)	0.279	(0.028)		
R013803b**	RH	3	1.200	(0.060)	0.498	(0.027)	0.000	(0.000)		
R013804b**	RH	4	0.524	(0.050)	-0.788	(0.174)	0.211	(0.047)		
R013805b**	RH	5	0.616	(0.027)	1.224	(0.046)	0.000	(0.000)	-0.642	(0.067)
									0.642	(0.084)
R013806b**	RH	6	0.688	(0.041)	0.316	(0.042)	0.000	(0.000)		
R013807b**	RH	7	1.070	(0.107)	0.616	(0.059)	0.219	(0.025)		
R013808b**	RH	8	0.857	(0.050)	0.593	(0.038)	0.000	(0.000)		
R013809b**	RH	9	0.773	(0.047)	0.501	(0.041)	0.000	(0.000)		
R013810b**	RH	10	0.616	(0.040)	-0.492	(0.060)	0.000	(0.000)		
R015501	RM	1	0.866	(0.060)	-0.529	(0.094)	0.323	(0.034)		
R015502	RM	2	0.823	(0.058)	0.265	(0.062)	0.194	(0.024)		

**Note*:* *a = item parameters are based on only 1994 data; b = item parameters are based on only 1992 data.*

***Note*: *This block name identifies two different sets of items for 1992 and 1994–a=1994 items; b=1992 items.*

Table D-7 (continued)

IRT Parameters for the Reading Main Samples
Reading to Gain Information Scale, Age 17/Grade 12

NAEP ID	Block	Item	A	S.E.	B	S.E.	C	S.E.	D	S.E.
R015503	RM	3	0.486	(0.037)	-2.440	(0.176)	0.000	(0.000)		
R015504	RM	4	0.855	(0.058)	0.015	(0.069)	0.245	(0.026)		
R015505	RM	5	0.623	(0.038)	-0.933	(0.071)	0.000	(0.000)		
R015506	RM	6	0.832	(0.063)	0.130	(0.074)	0.270	(0.027)		
R015507	RM	7	0.905	(0.023)	0.515	(0.016)	0.000	(0.000)	1.415	(0.032)
									-0.055	(0.027)
									-1.360	(0.048)
R015508	RM	8	1.004	(0.055)	-0.523	(0.061)	0.200	(0.027)		
R015509	RM	9	1.399	(0.065)	0.088	(0.024)	0.000	(0.000)		
R015510	RM	10	1.659	(0.103)	0.160	(0.035)	0.321	(0.018)		
R015511	RM	11	0.787	(0.056)	-0.533	(0.104)	0.297	(0.036)		
R015512	RM	12	1.031	(0.053)	-0.774	(0.045)	0.000	(0.000)		
R015513	RM	13	0.908	(0.052)	-0.700	(0.076)	0.219	(0.031)		
R015514	RM	14	0.582	(0.025)	0.406	(0.029)	0.000	(0.000)	1.221	(0.066)
									-0.095	(0.059)
									-1.126	(0.080)
R015515	RM	15	0.857	(0.067)	1.030	(0.048)	0.124	(0.015)		
R015516	RM	16	0.706	(0.054)	-0.059	(0.094)	0.250	(0.032)		
R015601b**	RN	1	0.496	(0.020)	-0.711	(0.035)	0.000	(0.000)	2.471	(0.130)
									-0.768	(0.063)
									-1.704	(0.071)
R015602b**	RN	2	0.857	(0.094)	0.012	(0.114)	0.355	(0.039)		
R015603b**	RN	3	0.563	(0.087)	0.438	(0.183)	0.370	(0.046)		
R015604b**	RN	4	1.146	(0.054)	-0.469	(0.034)	0.000	(0.000)		
R015605b**	RN	5	1.398	(0.137)	0.599	(0.049)	0.285	(0.022)		
R015606b**	RN	6	0.567	(0.108)	1.663	(0.168)	0.227	(0.032)		
R015607b**	RN	7	1.324	(0.061)	0.011	(0.025)	0.000	(0.000)		
R015608b**	RN	8	0.946	(0.115)	0.927	(0.070)	0.235	(0.025)		
R015609b**	RN	9	0.884	(0.032)	-0.030	(0.019)	0.000	(0.000)	0.953	(0.050)
									-0.578	(0.043)
									-0.375	(0.045)
R015610b**	RN	10	0.902	(0.069)	-0.294	(0.080)	0.183	(0.032)		
R015611b**	RN	11	0.798	(0.076)	-0.649	(0.139)	0.305	(0.047)		
R015612b**	RN	12	0.425	(0.021)	0.499	(0.036)	0.000	(0.000)	0.980	(0.082)
									-0.074	(0.082)
									-0.906	(0.102)
R016401a**	RH	1	0.818	(0.032)	-0.669	(0.032)	0.000	(0.000)	1.233	(0.063)
									-1.233	(0.039)
R016402a**	RH	2	0.346	(0.020)	-0.591	(0.059)	0.000	(0.000)	0.299	(0.105)
									-0.299	(0.090)

********Note****: a = item parameters are based on only 1994 data; b = item parameters are based on only 1992 data.*

*********Note****: This block name identifies two different sets of items for 1992 and 1994–a=1994 items; b=1992 items.*

Table D-7 (continued)

IRT Parameters for the Reading Main Samples
Reading to Gain Information Scale, Age 17/Grade 12

NAEP ID	Block	Item	A	S.E.	B	S.E.	C	S.E.	D	S.E.
R016403a**	RH	3	0.690	(0.029)	0.018	(0.033)	0.000	(0.000)	0.972	(0.052)
									-0.972	(0.050)
R016404a**	RH	4	0.475	(0.015)	0.232	(0.055)	0.000	(0.000)	2.201	(0.078)
									-2.201	(0.087)
R016405a**	RH	5	0.426	(0.015)	-0.736	(0.047)	0.000	(0.000)	-1.747	(0.120)
									1.747	(0.110)
R016406a**	RH	6	0.485	(0.053)	-0.722	(0.220)	0.255	(0.054)		
R016407a**	RH	7	0.513	(0.028)	0.124	(0.040)	0.000	(0.000)	0.599	(0.070)
									-0.599	(0.069)
R016408a**	RH	8	0.445	(0.022)	1.243	(0.045)	0.000	(0.000)	0.439	(0.083)
									0.871	(0.093)
									-1.311	(0.145)
R016501a**	RN	1	0.989	(0.030)	0.492	(0.033)	0.000	(0.000)	1.667	(0.041)
									-1.667	(0.061)
R016502a**	RN	2	0.891	(0.032)	1.164	(0.032)	0.000	(0.000)	1.453	(0.036)
									-1.453	(0.087)
R016601a**	RN	3	0.687	(0.027)	0.351	(0.034)	0.000	(0.000)	1.147	(0.049)
									-1.147	(0.056)
R016602a**	RN	4	0.592	(0.028)	0.499	(0.037)	0.000	(0.000)	0.843	(0.054)
									-0.843	(0.064)
R016604a**	RN	6	0.632	(0.023)	-0.574	(0.040)	0.000	(0.000)	1.589	(0.074)
									-1.589	(0.051)
R016605a**	RN	7	0.259	(0.014)	1.676	(0.111)	0.000	(0.000)	-1.677	(0.147)
									1.677	(0.180)
R016701a**	RN	8	0.818	(0.025)	0.737	(0.032)	0.000	(0.000)	2.272	(0.053)
									-0.788	(0.050)
									-1.484	(0.112)

***Note*:** *a = item parameters are based on only 1994 data; b = item parameters are based on only 1992 data.*

****Note*:** *This block name identifies two different sets of items for 1992 and 1994–a=1994 items; b=1992 items.*

Table D-8

IRT Parameters for the Reading Main Samples

Reading to Perform a Task Scale, Age 17/Grade 12

NAEP ID	Block	Item	A	S.E.	B	S.E.	C	S.E.	D	S.E.
R013301b**	RI	1	0.575	(0.061)	-0.822	(0.223)	0.318	(0.061)		
R013302b**	RI	2	1.043	(0.056)	0.909	(0.036)	0.000	(0.000)		
R013303b**	RI	3	0.995	(0.092)	-0.271	(0.108)	0.337	(0.042)		
R013304b**	RI	4	0.649	(0.043)	1.141	(0.065)	0.000	(0.000)		
R013305b**	RI	5	1.372	(0.129)	-1.841	(0.122)	0.282	(0.056)		
R013306b**	RI	6	0.746	(0.068)	-0.890	(0.163)	0.308	(0.055)		
R013307b**	RI	7	0.805	(0.042)	0.243	(0.036)	0.000	(0.000)		
R013308b**	RI	8	1.066	(0.096)	-0.281	(0.101)	0.349	(0.041)		
R013309b**	RI	9	0.794	(0.107)	0.742	(0.103)	0.313	(0.034)		
R013310b**	RI	10	0.649	(0.038)	0.011	(0.045)	0.000	(0.000)		
R013311b**	RI	11	1.016	(0.101)	-0.143	(0.109)	0.389	(0.040)		
R013312b**	RI	12	0.546	(0.029)	1.702	(0.059)	0.000	(0.000)	0.116	(0.064)
									0.239	(0.097)
									-0.355	(0.151)
R013401	RJ	1	0.933	(0.052)	-0.062	(0.054)	0.180	(0.023)		
R013402	RJ	2	0.882	(0.045)	-0.650	(0.045)	0.000	(0.000)		
R013403	RJ	3	0.446	(0.010)	-0.005	(0.025)	0.000	(0.000)	-2.281	(0.085)
									2.281	(0.085)
R013404	RJ	4	0.738	(0.051)	-0.950	(0.130)	0.311	(0.044)		
R013405	RJ	5	0.876	(0.032)	-0.919	(0.035)	0.000	(0.000)		
R013406	RJ	6	0.654	(0.027)	0.470	(0.033)	0.000	(0.000)		
R013407	RJ	7	0.595	(0.025)	-1.083	(0.052)	0.000	(0.000)		
R013408	RJ	8	0.710	(0.048)	-0.229	(0.091)	0.205	(0.032)		
R013409	RJ	9	0.797	(0.031)	-1.035	(0.041)	0.000	(0.000)		
R013410	RJ	10	0.729	(0.054)	-1.358	(0.160)	0.332	(0.053)		
R013411	RJ	11	0.558	(0.038)	0.242	(0.053)	0.000	(0.000)		
R013412	RJ	12	0.410	(0.042)	-2.642	(0.261)	0.000	(0.000)		
R013901	RK	1	1.600	(0.112)	0.378	(0.037)	0.366	(0.016)		
R013902	RK	2	0.743	(0.028)	-0.942	(0.039)	0.000	(0.000)		
R013903	RK	3	1.007	(0.034)	0.373	(0.022)	0.000	(0.000)		
R013904	RK	4	0.726	(0.029)	-1.485	(0.053)	0.000	(0.000)		
R013905	RK	5	0.998	(0.214)	2.562	(0.247)	0.250	(0.010)		
R013906	RK	6	0.527	(0.026)	0.910	(0.050)	0.000	(0.000)		
R013907	RK	7	1.024	(0.071)	0.287	(0.054)	0.273	(0.021)		
R013908	RK	8	0.689	(0.040)	-1.095	(0.072)	0.000	(0.000)		
R013909	RK	9	0.722	(0.050)	-0.070	(0.086)	0.218	(0.030)		
R013910	RK	10	1.001	(0.049)	0.130	(0.032)	0.000	(0.000)		
R013911	RK	11	0.759	(0.054)	-0.954	(0.133)	0.335	(0.044)		
R013912	RK	12	0.508	(0.072)	1.179	(0.126)	0.281	(0.032)		
R013913	RK	13	0.494	(0.034)	-0.138	(0.061)	0.000	(0.000)		

**Note:* *a = item parameters are based on only 1994 data; b = item parameters are based on only 1992 data.*

***Note:* *This block name identifies two different sets of items for 1992 and 1994–a=1994 items; b=1992 items.*

Table D-8 (continued)

IRT Parameters for the Reading Main Samples
Reading to Perform a Task Scale, Age 17/Grade 12

NAEP ID	Block	Item	A	S.E.	B	S.E.	C	S.E.	D	S.E.
R013914	RK	14	0.364	(0.040)	0.116	(0.217)	0.342	(0.040)		
R013915	RK	15	0.523	(0.022)	1.494	(0.054)	0.000	(0.000)	-2.253	(0.137)
									2.253	(0.150)
R016101a**	RI	1	0.583	(0.025)	-0.503	(0.035)	0.000	(0.000)	0.074	(0.066)
									-0.074	(0.056)
R016102a**	RI	2	1.106	(0.086)	-1.418	(0.102)	0.256	(0.043)		
R016103a**	RI	3	1.071	(0.087)	-0.554	(0.088)	0.306	(0.036)		
R016104a**	RI	4	0.727	(0.044)	-2.007	(0.094)	0.000	(0.000)		
R016105a**	RI	5	1.233	(0.103)	-1.368	(0.101)	0.303	(0.044)		
R016106a**	RI	6	0.910	(0.075)	-0.265	(0.088)	0.234	(0.034)		
R016107a**	RI	7	0.654	(0.025)	-0.886	(0.037)	0.000	(0.000)	-0.300	(0.071)
									0.300	(0.057)
R016108a**	RI	8	0.392	(0.017)	-0.520	(0.047)	0.000	(0.000)	-0.676	(0.099)
									0.676	(0.089)
R016109a**	RI	9	0.380	(0.015)	0.127	(0.058)	0.000	(0.000)	1.884	(0.089)
									-1.884	(0.093)

***Note*:** *a = item parameters are based on only 1994 data; b = item parameters are based on only 1992 data.*

****Note*:** *This block name identifies two different sets of items for 1992 and 1994–a=1994 items; b=1992 items.*

Table D-9

IRT Parameters for the U.S. History Main Sample
Democracy Scale, Age 9/Grade 4

NAEP ID	Block	Item	A	S.E.	B	S.E.	C	S.E.	D	S.E.
H028501	HC	5	1.206	(0.158)	1.301	(0.066)	0.203	(0.016)		
H028601	HC	6	2.678	(0.150)	1.283	(0.033)	0.141	(0.009)		
H028602	HC	7	2.787	(0.188)	1.303	(0.040)	0.294	(0.011)		
H028901	HC	11	1.333	(0.127)	0.577	(0.047)	0.261	(0.021)		
H029101	HC	14	0.696	(0.165)	1.635	(0.181)	0.355	(0.029)		
H029901	HD	8	2.013	(0.148)	1.288	(0.043)	0.177	(0.011)		
H030601	HE	3	1.127	(0.191)	1.657	(0.121)	0.248	(0.015)		
H030701	HE	7	0.553	(0.022)	0.785	(0.030)	0.000	(0.000)	1.273	(0.058)
									-0.038	(0.059)
									-1.235	(0.095)
H030801	HE	8	0.802	(0.127)	1.188	(0.091)	0.265	(0.025)		
H031001	HE	10	1.655	(0.159)	0.678	(0.038)	0.271	(0.018)		
H031301	HE	13	0.702	(0.088)	0.338	(0.122)	0.307	(0.037)		
H031401	HE	14	1.533	(0.121)	-0.703	(0.068)	0.325	(0.033)		
H031402	HE	15	0.691	(0.031)	0.349	(0.030)	0.000	(0.000)	0.692	(0.047)
									-0.692	(0.050)
H032001	HF	7	0.673	(0.079)	0.764	(0.087)	0.182	(0.029)		
H032002	HF	8	1.617	(0.192)	1.348	(0.073)	0.407	(0.014)		
H032501	HF	13	1.270	(0.127)	0.816	(0.043)	0.167	(0.018)		
H032502	HF	14	0.842	(0.082)	0.336	(0.078)	0.204	(0.029)		
H032503	HF	15	1.144	(0.065)	1.181	(0.034)	0.000	(0.000)	0.107	(0.036)
									-0.107	(0.055)
H032901	HG	4	1.554	(0.139)	0.785	(0.033)	0.160	(0.015)		
H033401	HG	9	2.269	(0.150)	1.153	(0.035)	0.215	(0.011)		
H033601	HG	11	1.984	(0.143)	0.965	(0.032)	0.221	(0.013)		
H033602	HG	12	0.931	(0.111)	0.636	(0.074)	0.272	(0.028)		
H033701	HG	13	1.170	(0.095)	-0.023	(0.066)	0.263	(0.029)		
H033801	HG	14	0.763	(0.102)	0.774	(0.090)	0.246	(0.031)		
H034701	HH	7	0.355	(0.057)	0.591	(0.257)	0.255	(0.051)		
H034702	HH	8	0.631	(0.071)	3.139	(0.258)	0.000	(0.000)	0.361	(0.079)
									-0.361	(0.326)
H034801	HH	9	1.451	(0.149)	1.147	(0.049)	0.216	(0.014)		
H034901	HH	10	0.645	(0.060)	-0.656	(0.153)	0.245	(0.046)		
H035301	HH	15	1.100	(0.188)	1.423	(0.094)	0.214	(0.018)		
H035401	HH	16	0.807	(0.143)	1.189	(0.103)	0.295	(0.028)		

Table D-10

IRT Parameters for the U.S. History Main Sample
Cultures Scale, Age 9/Grade 4

NAEP ID	Block	Item	A	S.E.	B	S.E.	C	S.E.	D	S.E.
H028301	HC	2	1.245	(0.489)	2.659	(0.616)	0.197	(0.011)		
H028302	HC	3	1.408	(0.111)	0.003	(0.051)	0.314	(0.024)		
H028401	HC	4	0.535	(0.108)	1.190	(0.161)	0.345	(0.036)		
H029001	HC	12	2.737	(0.161)	1.140	(0.031)	0.213	(0.011)		
H029002	HC	13	0.699	(0.033)	0.474	(0.029)	0.000	(0.000)	0.581 -0.581	(0.045) (0.050)
H029401	HD	2	1.275	(0.103)	0.098	(0.053)	0.267	(0.025)		
H029601	HD	4	0.937	(0.116)	0.937	(0.067)	0.220	(0.023)		
H029701	HD	5	0.751	(0.027)	-0.375	(0.030)	0.000	(0.000)	1.158 -1.158	(0.057) (0.039)
H029702	HD	6	0.790	(0.031)	0.102	(0.027)	0.000	(0.000)	0.819 -0.819	(0.044) (0.040)
H030001	HD	9	2.017	(0.181)	0.722	(0.032)	0.275	(0.017)		
H030002	HD	10	1.586	(0.157)	0.807	(0.038)	0.209	(0.018)		
H030201	HD	13	0.715	(0.035)	1.845	(0.045)	0.000	(0.000)	1.296 -1.296	(0.042) (0.141)
H030602	HE	4	1.354	(0.135)	0.064	(0.071)	0.482	(0.027)		
H030603	HE	5	1.079	(0.040)	0.339	(0.017)	0.000	(0.000)	-0.018 0.018	(0.033) (0.032)
H030604	HE	6	0.937	(0.031)	-0.200	(0.026)	0.000	(0.000)	1.126 -1.126	(0.047) (0.033)
H031101	HE	11	0.839	(0.038)	0.916	(0.029)	0.000	(0.000)	0.639 -0.639	(0.035) (0.052)
H031201	HE	12	1.719	(0.152)	1.209	(0.047)	0.189	(0.013)		
H031701	HF	2	0.586	(0.018)	1.187	(0.026)	0.000	(0.000)	-0.501 1.838 -1.337	(0.071) (0.074) (0.100)
H031901	HF	5	1.508	(0.191)	1.501	(0.085)	0.252	(0.013)		
H031902	HF	6	2.363	(0.166)	0.820	(0.029)	0.324	(0.014)		
H032601	HG	1	0.864	(0.041)	0.895	(0.029)	0.000	(0.000)	0.195 -0.195	(0.036) (0.049)
H032801	HG	3	1.245	(0.338)	1.973	(0.254)	0.349	(0.015)		
H033001	HG	5	0.846	(0.048)	1.480	(0.048)	0.000	(0.000)	0.498 -0.498	(0.036) (0.078)
H033101	HG	6	0.658	(0.078)	-0.865	(0.223)	0.394	(0.054)		
H033201	HG	7	0.509	(0.029)	1.262	(0.056)	0.000	(0.000)	0.647 -0.647	(0.056) (0.087)
H033501	HG	10	1.916	(0.168)	0.847	(0.035)	0.283	(0.016)		
H034201	HH	2	1.355	(0.207)	1.339	(0.085)	0.365	(0.016)		
H034301	HH	3	1.714	(0.222)	1.548	(0.092)	0.298	(0.012)		
H034601	HH	6	1.339	(0.166)	0.882	(0.052)	0.320	(0.020)		

BEST COPY AVAILABLE

Table D-11

IRT Parameters for the U.S. History Main Sample
Technology Scale, Age 9/Grade 4

NAEP ID	Block	Item	A S.E.	B S.E.	C S.E.	D S.E.
H028701	HC	8	0.776 (0.028)	0.131 (0.028)	0.000 (0.000)	0.962 (0.044)
						-0.962 (0.041)
H028702	HC	9	1.139 (0.041)	0.595 (0.021)	0.000 (0.000)	0.708 (0.029)
						-0.708 (0.035)
H028801	HC	10	0.680 (0.038)	1.476 (0.053)	0.000 (0.000)	0.434 (0.045)
						-0.434 (0.080)
H029201	HC	15	1.252 (0.155)	0.787 (0.057)	0.329 (0.023)	
H029301	HD	1	1.644 (0.142)	1.142 (0.044)	0.248 (0.014)	
H029501	HD	3	0.525 (0.017)	0.172 (0.024)	0.000 (0.000)	-0.906 (0.088)
						0.937 (0.090)
						-0.031 (0.065)
H029801	HD	7	0.516 (0.070)	0.067 (0.206)	0.291 (0.051)	
H030101	HD	11	0.996 (0.046)	0.842 (0.024)	0.000 (0.000)	0.225 (0.034)
						-0.225 (0.043)
H030102	HD	12	1.430 (0.139)	0.541 (0.046)	0.287 (0.021)	
H030401	HE	1	2.452 (0.145)	0.842 (0.025)	0.245 (0.013)	
H030901	HE	9	1.406 (0.133)	0.457 (0.048)	0.316 (0.021)	
H031601	HF	1	2.311 (0.194)	1.580 (0.060)	0.217 (0.009)	
H031801	HF	3	0.596 (0.022)	0.795 (0.031)	0.000 (0.000)	-0.805 (0.064)
						0.805 (0.070)
H031802	HF	4	0.867 (0.035)	0.554 (0.022)	0.000 (0.000)	0.154 (0.038)
						-0.154 (0.041)
H032101	HF	9	0.831 (0.060)	-1.882 (0.149)	0.272 (0.059)	
H032301	HF	11	0.863 (0.040)	0.927 (0.029)	0.000 (0.000)	0.147 (0.038)
						-0.147 (0.050)
H032401	HF	12	0.493 (0.045)	-2.110 (0.277)	0.303 (0.067)	
H033901	HG	15	1.074 (0.202)	1.609 (0.126)	0.263 (0.018)	
H034401	HH	4	0.736 (0.030)	0.515 (0.028)	0.000 (0.000)	0.696 (0.042)
						-0.696 (0.048)
H035001	HH	11	1.016 (0.053)	1.319 (0.039)	0.000 (0.000)	-0.012 (0.038)
						0.012 (0.060)
H035002	HH	12	1.962 (0.165)	0.746 (0.033)	0.300 (0.016)	
H035101	HH	13	0.857 (0.035)	0.115 (0.026)	0.000 (0.000)	0.663 (0.043)
						-0.663 (0.039)

Table D-12

IRT Parameters for the U.S. History Main Sample
World Role Scale, Age 9/Grade 4

NAEP ID	Block	Item	A	S.E.	B	S.E.	C	S.E.	D	S.E.
H028201	HC	1	0.526	(0.024)	1.712	(0.052)	0.000	(0.000)	0.955	(0.054)
									-0.704	(0.100)
									-0.251	(0.157)
H030301	HD	14	0.734	(0.076)	0.447	(0.103)	0.190	(0.035)		
H030302	HD	15	0.385	(0.023)	2.553	(0.147)	0.000	(0.000)	-1.550	(0.150)
									1.550	(0.212)
H030501	HE	2	0.813	(0.087)	0.907	(0.078)	0.208	(0.025)		
H031501	HE	16	1.498	(0.109)	0.119	(0.045)	0.177	(0.022)		
H032201	HF	10	0.571	(0.069)	0.934	(0.122)	0.183	(0.035)		
H032701	HG	2	0.895	(0.032)	1.233	(0.026)	0.000	(0.000)	0.004	(0.045)
									0.254	(0.061)
									-0.258	(0.071)
H033301	HG	8	0.615	(0.028)	1.979	(0.070)	0.000	(0.000)	-0.923	(0.089)
									0.923	(0.122)
H034001	HG	16	0.526	(0.108)	2.266	(0.219)	0.178	(0.028)		
H034101	HH	1	0.760	(0.023)	0.832	(0.036)	0.000	(0.000)	1.895	(0.041)
									-1.895	(0.078)
H034501	HH	5	0.423	(0.062)	4.697	(0.622)	0.000	(0.000)		
H035201	HH	14	0.644	(0.220)	3.421	(0.640)	0.252	(0.017)		
H035501	HH	17	0.741	(0.098)	1.237	(0.095)	0.187	(0.026)		

Table D-13

IRT Parameters for the U.S. History Main Sample
Democracy Scale, Age 13/Grade 8

NAEP ID	Block	Item	A	S.E.	B	S.E.	C	S.E.	D	S.E.
H032901	HG	4	1.069	(0.073)	-0.371	(0.067)	0.172	(0.032)		
H033401	HG	9	1.132	(0.095)	0.192	(0.068)	0.270	(0.027)		
H033601	HG	11	1.464	(0.107)	0.235	(0.045)	0.208	(0.020)		
H033602	HG	12	1.294	(0.099)	0.092	(0.056)	0.242	(0.025)		
H033701	HG	13	0.460	(0.041)	-1.951	(0.223)	0.202	(0.055)		
H033801	HG	14	0.441	(0.043)	-0.356	(0.199)	0.196	(0.050)		
H034701	HH	7	0.590	(0.049)	-1.009	(0.173)	0.223	(0.056)		
H034702	HH	8	0.785	(0.041)	2.041	(0.061)	0.000	(0.000)	0.613 -0.613	(0.045) (0.110)
H034801	HH	9	1.024	(0.083)	-0.127	(0.085)	0.267	(0.035)		
H034901	HH	10	0.519	(0.046)	-2.040	(0.214)	0.211	(0.057)		
H035301	HH	15	0.345	(0.039)	-0.005	(0.247)	0.197	(0.051)		
H035401	HH	16	0.622	(0.050)	-1.087	(0.163)	0.219	(0.055)		
H035701	HC	3	0.848	(0.157)	1.947	(0.127)	0.230	(0.018)		
H035801	HC	4	0.620	(0.034)	2.234	(0.089)	0.000	(0.000)	-0.384 0.384	(0.075) (0.126)
H036601	HC	14	0.856	(0.099)	1.437	(0.073)	0.136	(0.018)		
H037101	HD	3	0.922	(0.033)	0.358	(0.026)	0.000	(0.000)	0.983 -0.983	(0.035) (0.042)
H037201	HD	4	0.380	(0.042)	0.009	(0.230)	0.197	(0.051)		
H037202	HD	5	0.719	(0.056)	-0.413	(0.116)	0.193	(0.043)		
H037203	HD	6	0.876	(0.043)	2.156	(0.052)	0.000	(0.000)	0.896 -0.896	(0.040) (0.129)
H037302	HD	8	1.232	(0.136)	1.302	(0.058)	0.245	(0.015)		
H037401	HD	9	0.992	(0.039)	1.500	(0.029)	0.000	(0.000)	0.368 -0.613 0.245	(0.038) (0.081) (0.096)
H037601	HD	14	0.663	(0.036)	1.976	(0.071)	0.000	(0.000)	0.233 -0.233	(0.055) (0.106)
H037901	HE	1	0.509	(0.063)	0.770	(0.153)	0.200	(0.041)		
H038001	HE	2	0.877	(0.068)	-1.720	(0.131)	0.219	(0.057)		
H038101	HE	3	0.745	(0.099)	0.924	(0.108)	0.300	(0.031)		
H038102	HE	4	1.614	(0.141)	0.904	(0.040)	0.200	(0.014)		
H038103	HE	5	0.562	(0.026)	2.241	(0.075)	0.000	(0.000)	-0.912 0.549 0.363	(0.096) (0.159) (0.208)
H038201	HE	6	1.321	(0.118)	0.981	(0.046)	0.172	(0.015)		
H038202	HE	7	0.793	(0.105)	1.188	(0.088)	0.244	(0.025)		
H039001	HE	17	0.801	(0.034)	1.114	(0.035)	0.000	(0.000)	0.113 -0.113	(0.046) (0.062)
H039201	HF	2	0.666	(0.062)	0.474	(0.102)	0.165	(0.033)		
H039501	HF	5	0.561	(0.065)	0.565	(0.144)	0.208	(0.041)		
H040001	HF	10	1.129	(0.042)	0.963	(0.025)	0.000	(0.000)	0.912 -0.912	(0.028) (0.048)

Table D-13 (continued)

IRT Parameters for the U.S. History Main Sample
Democracy Scale, Age 13/Grade 8

NAEP ID	Block	Item	A	S.E.	B	S.E.	C	S.E.	D	S.E.
H040301	HF	15	0.592	(0.113)	1.922	(0.156)	0.221	(0.028)		
H040401	HF	16	0.705	(0.056)	-0.868	(0.139)	0.219	(0.052)		
H040701	HI	2	0.642	(0.077)	1.157	(0.097)	0.161	(0.028)		
H040803	HI	5	0.619	(0.062)	0.398	(0.122)	0.191	(0.038)		
H040901	HI	6	0.636	(0.130)	2.282	(0.200)	0.168	(0.021)		
H041801	HI	16	0.358	(0.015)	1.719	(0.076)	0.000	(0.000)	-1.829	(0.121)
									1.829	(0.146)
H041802	HI	17	0.724	(0.049)	2.408	(0.104)	0.000	(0.000)	0.116	(0.065)
									-0.116	(0.147)
H042001	HJ	2	0.789	(0.133)	1.765	(0.114)	0.226	(0.020)		
H042501	HJ	8	1.464	(0.120)	1.697	(0.061)	0.183	(0.010)		
H042601	HJ	9	0.642	(0.057)	-0.150	(0.133)	0.204	(0.044)		
H042602	HJ	10	1.345	(0.147)	2.323	(0.122)	0.089	(0.007)		
H042801	HJ	13	0.603	(0.023)	1.626	(0.040)	0.000	(0.000)	1.662	(0.046)
							-1.202	(0.098)	-0.460	(0.170)
H053901	HD	10	1.142	(0.055)	2.015	(0.042)	0.000	(0.000)	0.746	(0.035)
									-0.746	(0.106)

Table D-14

IRT Parameters for the U.S. History Main Sample
Cultures Scale, Age 13/Grade 8

NAEP ID	Block	Item	A	S.E.	B	S.E.	C	S.E.	D	S.E.
H032601	HG	1	0.392	(0.021)	-0.750	(0.054)	0.000	(0.000)	0.551 -0.551	(0.092) (0.073)
H032801	HG	3	0.481	(0.071)	1.095	(0.150)	0.195	(0.038)		
H033001	HG	5	0.556	(0.025)	0.732	(0.039)	0.000	(0.000)	0.930 -0.930	(0.052) (0.069)
H033101	HG	6	0.618	(0.050)	-1.538	(0.183)	0.228	(0.058)		
H033201	HG	7	0.379	(0.017)	0.062	(0.050)	0.000	(0.000)	1.367 -1.367	(0.082) (0.083)
H033501	HG	10	0.914	(0.082)	-0.260	(0.101)	0.283	(0.038)		
H034201	HH	2	0.619	(0.078)	0.515	(0.133)	0.267	(0.039)		
H034301	HH	3	0.547	(0.057)	-0.203	(0.174)	0.228	(0.049)		
H034601	HH	6	0.461	(0.042)	-1.175	(0.224)	0.220	(0.057)		
H035601	HC	1	1.403	(0.098)	-0.230	(0.052)	0.213	(0.025)		
H035602	HC	2	1.300	(0.111)	0.725	(0.045)	0.212	(0.018)		
H036201	HC	9	0.496	(0.048)	-0.783	(0.213)	0.230	(0.057)		
H036702	HC	16	1.124	(0.083)	-1.736	(0.108)	0.215	(0.052)		
H036901	HD	1	1.194	(0.113)	0.731	(0.053)	0.250	(0.020)		
H037301	HD	7	1.235	(0.093)	0.282	(0.047)	0.190	(0.021)		
H037801	HD	16	0.533	(0.046)	-1.330	(0.201)	0.223	(0.057)		
H037802	HD	17	1.107	(0.232)	2.683	(0.277)	0.119	(0.009)		
H038501	HE	10	1.132	(0.103)	0.874	(0.050)	0.173	(0.018)		
H038601	HE	11	0.413	(0.033)	2.730	(0.177)	0.000	(0.000)	0.218 -0.218	(0.084) (0.184)
H038801	HE	14	0.747	(0.083)	0.656	(0.090)	0.219	(0.030)		
H038802	HE	15	0.698	(0.064)	-0.301	(0.130)	0.226	(0.044)		
H038901	HE	16	0.770	(0.094)	1.004	(0.083)	0.208	(0.026)		
H039401	HF	4	0.715	(0.024)	0.439	(0.033)	0.000	(0.000)	1.332 -1.332	(0.044) (0.055)
H039801	HF	8	1.737	(0.116)	0.224	(0.032)	0.163	(0.017)		
H040601	HI	1	0.544	(0.052)	-0.380	(0.165)	0.207	(0.048)		
H040801	HI	3	0.863	(0.070)	0.140	(0.073)	0.173	(0.028)		
H040802	HI	4	0.664	(0.052)	-0.615	(0.127)	0.185	(0.043)		
H041401	HI	12	0.614	(0.025)	0.934	(0.031)	0.000	(0.000)	1.005 -0.292 -0.713	(0.049) (0.062) (0.091)
H043201	HK	1	0.679	(0.026)	0.093	(0.030)	0.000	(0.000)	0.994 -0.994	(0.048) (0.048)
H043301	HK	2	0.901	(0.080)	0.655	(0.060)	0.155	(0.022)		
H043401	HK	3	0.705	(0.029)	0.250	(0.027)	0.000	(0.000)	0.352 -0.352	(0.045) (0.047)
H043501	HK	4	0.545	(0.024)	1.304	(0.038)	0.000	(0.000)	0.568 0.435 -1.003	(0.057) (0.071) (0.116)

Table D-14 (continued)

IRT Parameters for the U.S. History Main Sample
Cultures Scale, Age 13/Grade 8

NAEP ID	Block	Item	A	S.E.	B	S.E.	C	S.E.	D	S.E.
H043601	HK	5	0.639	(0.025)	0.768	(0.032)	0.000	(0.000)	-0.237	(0.052)
									0.237	(0.062)
H043701	HK	6	0.447	(0.017)	0.503	(0.037)	0.000	(0.000)	-0.771	(0.077)
									0.771	(0.082)
H043702	HK	7	1.706	(0.129)	1.326	(0.044)	0.181	(0.011)		
H043703	HK	8	1.125	(0.082)	0.373	(0.046)	0.138	(0.020)		
H043704	HK	9	1.218	(0.164)	1.953	(0.116)	0.223	(0.012)		
H043705	HK	10	0.692	(0.030)	0.881	(0.034)	0.000	(0.000)	0.657	(0.042)
									-0.657	(0.059)
H043801	HK	11	0.896	(0.069)	-0.328	(0.088)	0.202	(0.035)		
H043901	HK	12	1.008	(0.085)	-0.334	(0.091)	0.284	(0.036)		
H044001	HK	13	0.739	(0.032)	1.501	(0.035)	0.000	(0.000)	0.858	(0.041)
									0.074	(0.061)
									-0.932	(0.123)

Table D-15

IRT Parameters for the U.S. History Main Sample
Technology Scale, Age 13/Grade 8

NAEP ID	Block	Item	A	S.E.	B	S.E.	C	S.E.	D	S.E.
H033901	HG	15	1.040	(0.191)	1.326	(0.098)	0.376	(0.021)		
H034401	HH	4	0.553	(0.023)	-1.018	(0.042)	0.000	(0.000)	1.334 -1.334	(0.086) (0.049)
H035001	HH	11	0.851	(0.033)	0.186	(0.022)	0.000	(0.000)	0.117 -0.117	(0.040) (0.039)
H035002	HH	12	1.724	(0.130)	-0.349	(0.053)	0.333	(0.028)		
H035101	HH	13	0.698	(0.031)	-1.410	(0.043)	0.000	(0.000)	1.043 -1.043	(0.088) (0.041)
H035901	HC	5	0.668	(0.029)	0.059	(0.028)	0.000	(0.000)	0.365 -0.365	(0.049) (0.046)
H035902	HC	6	0.880	(0.044)	1.280	(0.040)	0.000	(0.000) -0.139	0.139 (0.062)	(0.038)
H036301	HC	10	1.200	(0.101)	-0.574	(0.091)	0.395	(0.036)		
H036401	HC	11	0.869	(0.082)	0.253	(0.080)	0.239	(0.030)		
H036402	HC	12	0.777	(0.030)	1.045	(0.026)	0.000	(0.000)	1.218 0.008 -1.225	(0.040) (0.046) (0.099)
H036701	HC	15	0.622	(0.062)	-1.096	(0.203)	0.326	(0.055)		
H036801	HC	17	0.566	(0.077)	0.571	(0.140)	0.267	(0.039)		
H037001	HD	2	0.958	(0.139)	1.002	(0.077)	0.325	(0.025)		
H037502	HD	13	1.204	(0.112)	0.230	(0.065)	0.336	(0.027)		
H038401	HE	9	1.614	(0.157)	1.278	(0.054)	0.256	(0.013)		
H039901	HF	9	0.639	(0.018)	0.398	(0.029)	0.000	(0.000)	1.958 -1.453 -0.504	(0.058) (0.063) (0.085)
H040101	HF	11	1.495	(0.224)	1.787	(0.122)	0.283	(0.012)		
H040102	HF	12	1.175	(0.099)	-0.244	(0.079)	0.354	(0.032)		
H040103	HF	13	1.089	(0.055)	1.320	(0.037)	0.000	(0.000)	0.079 -0.079	(0.035) (0.058)
H040201	HF	14	0.615	(0.025)	0.659	(0.030)	0.000	(0.000)	-0.407 0.407	(0.056) (0.063)
H040501	HF	17	0.554	(0.070)	0.520	(0.132)	0.237	(0.037)		
H041101	HI	8	1.047	(0.169)	1.536	(0.098)	0.208	(0.017)		
H041102	HI	9	0.971	(0.156)	1.208	(0.085)	0.333	(0.023)		
H041301	HI	11	0.627	(0.041)	1.990	(0.085)	0.000	(0.000)	0.507 -0.507	(0.051) (0.119)
H041501	HI	13	0.660	(0.070)	-0.543	(0.173)	0.345	(0.048)		
H042101	HJ	3	0.491	(0.062)	-0.009	(0.199)	0.290	(0.049)		
H042102	HJ	4	1.050	(0.163)	1.229	(0.078)	0.303	(0.021)		
H042701	HJ	11	1.099	(0.085)	-0.159	(0.070)	0.268	(0.030)		
H042702	HJ	12	1.269	(0.087)	-0.400	(0.060)	0.231	(0.029)		
H043001	HJ	16	0.385	(0.027)	2.211	(0.116)	0.000	(0.000)	0.838 -0.838	(0.074) (0.151)
H043101	HJ	17	0.356	(0.019)	1.265	(0.066)	0.000	(0.000)	1.349 -1.349	(0.080) (0.119)
H054001	HD	11	1.029	(0.098)	0.655	(0.054)	0.188	(0.022)		

Table D-16

IRT Parameters for the U.S. History Main Sample
World Role Scale, Age 13/Grade 8

NAEP ID	Block	Item	A	S.E.	B	S.E.	C	S.E.	D	S.E.
H032701	HG	2	0.497	(0.019)	0.129	(0.028)	0.000	(0.000)	0.222	(0.075)
									0.410	(0.075)
									-0.632	(0.073)
H033301	HG	8	0.611	(0.023)	0.727	(0.033)	0.000	(0.000)	-0.368	(0.058)
									0.368	(0.067)
H034001	HG	16	0.651	(0.124)	1.814	(0.154)	0.254	(0.026)		
H034101	HH	1	0.972	(0.029)	0.211	(0.025)	0.000	(0.000)	2.428	(0.058)
									-0.644	(0.034)
									-1.784	(0.065)
H034501	HH	5	1.095	(0.053)	1.031	(0.038)	0.000	(0.000)		
H035201	HH	14	1.018	(0.091)	0.352	(0.071)	0.261	(0.027)		
H035501	HH	17	0.916	(0.075)	-0.702	(0.117)	0.286	(0.047)		
H036001	HC	7	1.985	(0.182)	0.659	(0.038)	0.332	(0.016)		
H036101	HC	8	0.482	(0.015)	-0.156	(0.033)	0.000	(0.000)	-1.481	(0.090)
									1.481	(0.088)
H036501	HC	13	1.254	(0.111)	0.486	(0.055)	0.279	(0.022)		
H037501	HD	12	1.091	(0.107)	0.815	(0.059)	0.243	(0.020)		
H037701	HD	15	1.120	(0.136)	1.254	(0.065)	0.257	(0.017)		
H038301	HE	8	0.893	(0.032)	0.867	(0.026)	0.000	(0.000)	-0.213	(0.043)
									0.213	(0.052)
H038701	HE	12	0.586	(0.055)	-0.546	(0.179)	0.267	(0.052)		
H038702	HE	13	0.686	(0.034)	1.834	(0.054)	0.000	(0.000)	0.896	(0.045)
									-0.896	(0.107)
H039101	HF	1	0.756	(0.100)	0.589	(0.125)	0.395	(0.034)		
H039301	HF	3	1.591	(0.142)	1.408	(0.056)	0.325	(0.013)		
H039601	HF	6	1.501	(0.142)	1.642	(0.069)	0.291	(0.012)		
H039701	HF	7	1.296	(0.141)	1.579	(0.073)	0.287	(0.013)		
H041001	HI	7	1.652	(0.180)	2.058	(0.102)	0.265	(0.010)		
H041201	HI	10	0.758	(0.108)	1.441	(0.097)	0.232	(0.023)		
H041601	HI	14	1.311	(0.130)	1.724	(0.073)	0.181	(0.011)		
H041701	HI	15	1.923	(0.207)	2.295	(0.113)	0.143	(0.008)		
H041901	HJ	1	1.478	(0.154)	1.711	(0.079)	0.328	(0.012)		
H042201	HJ	5	0.777	(0.060)	2.704	(0.133)	0.000	(0.000)	0.260	(0.065)
									-0.260	(0.200)
H042301	HJ	6	0.878	(0.208)	2.417	(0.238)	0.217	(0.015)		
H042401	HJ	7	0.704	(0.175)	2.502	(0.284)	0.251	(0.019)		
H042901	HJ	14	1.954	(0.132)	-0.031	(0.034)	0.196	(0.019)		
H042902	HJ	15	1.054	(0.042)	0.857	(0.024)	0.000	(0.000)	0.201	(0.033)
									-0.201	(0.043)

BEST COPY AVAILABLE

Table D-17

IRT Parameters for the U.S. History Main Sample
Democracy Scale, Age 17/Grade 12

NAEP ID	Block	Item	A	S.E.	B	S.E.	C	S.E.	D	S.E.
H040701	HI	2	0.637	(0.050)	0.081	(0.097)	0.128	(0.032)		
H040803	HI	5	0.918	(0.069)	-0.237	(0.083)	0.178	(0.035)		
H040901	HI	6	0.683	(0.073)	1.174	(0.085)	0.117	(0.023)		
H041801	HI	16	0.494	(0.017)	0.509	(0.036)	0.000	(0.000)	-1.179	(0.084)
									1.179	(0.090)
H041802	HI	17	0.708	(0.033)	1.164	(0.041)	0.000	(0.000)	0.117	(0.050)
									-0.117	(0.070)
H042001	HJ	2	0.659	(0.074)	0.771	(0.103)	0.175	(0.032)		
H042501	HJ	8	1.197	(0.091)	0.517	(0.046)	0.136	(0.018)		
H042601	HJ	9	0.845	(0.065)	-0.658	(0.109)	0.202	(0.045)		
H042602	HJ	10	0.929	(0.111)	1.787	(0.092)	0.084	(0.012)		
H042801	HJ	13	0.562	(0.022)	0.652	(0.032)	0.000	(0.000)	1.451	(0.058)
									-0.795	(0.070)
									-0.656	(0.096)
H044401	HC	4	0.859	(0.064)	-0.471	(0.094)	0.177	(0.039)		
H044501	HC	5	0.537	(0.018)	0.896	(0.031)	0.000	(0.000)	-1.022	(0.091)
									1.689	(0.098)
									-0.667	(0.088)
H044901	HC	10	0.638	(0.082)	0.750	(0.124)	0.235	(0.037)		
H045701	HD	3	0.706	(0.071)	0.583	(0.093)	0.167	(0.031)		
H045901	HD	5	0.900	(0.034)	1.123	(0.026)	0.000	(0.000)	1.325	(0.035)
									-0.318	(0.050)
									-1.007	(0.092)
H046001	HD	6	0.642	(0.038)	1.171	(0.068)	0.000	(0.000)		
H046501	HD	12	1.655	(0.137)	1.447	(0.053)	0.201	(0.011)		
H046901	HD	17	1.484	(0.127)	0.693	(0.042)	0.164	(0.016)		
H046902	HD	18	1.202	(0.136)	0.841	(0.063)	0.282	(0.021)		
H047101	HE	2	0.739	(0.056)	-1.062	(0.133)	0.198	(0.052)		
H047201	HE	3	0.803	(0.029)	0.736	(0.022)	0.000	(0.000)	0.348	(0.043)
									-0.159	(0.057)
									-0.189	(0.063)
H047301	HE	4	0.549	(0.021)	1.666	(0.045)	0.000	(0.000)	-0.572	(0.080)
									1.216	(0.100)
									-0.644	(0.131)
H048601	HF	1	0.620	(0.057)	-0.347	(0.149)	0.200	(0.049)		
H048701	HF	2	0.656	(0.074)	0.781	(0.104)	0.175	(0.033)		
H048801	HF	3	1.447	(0.135)	1.592	(0.064)	0.191	(0.011)		
H048901	HF	4	0.532	(0.021)	1.649	(0.058)	0.000	(0.000)	-1.977	(0.127)
									1.977	(0.145)
H049101	HF	6	0.832	(0.084)	0.667	(0.079)	0.185	(0.028)		
H049102	HF	7	0.523	(0.044)	-1.254	(0.178)	0.187	(0.052)		
H049401	HF	10	0.641	(0.038)	2.428	(0.089)	0.000	(0.000)	0.318	(0.065)
									-0.014	(0.141)
									-0.303	(0.266)

Table D-17 (continued)

IRT Parameters for the U.S. History Main Sample
Democracy Scale, Age 17/Grade 12

NAEP ID	Block	Item	A	S.E.	B	S.E.	C	S.E.	D	S.E.
H050101	HG	2	0.701	(0.025)	1.204	(0.031)	0.000	(0.000)	-0.011	(0.052)
									-0.651	(0.099)
									0.661	(0.107)
H050401	HG	7	0.669	(0.096)	1.498	(0.109)	0.190	(0.026)		
H050501	HG	8	0.897	(0.098)	1.094	(0.069)	0.174	(0.021)		
H051401	HH	2	0.672	(0.138)	2.024	(0.179)	0.278	(0.024)		
H051501	HH	3	2.019	(0.154)	1.152	(0.041)	0.289	(0.012)		
H051601	HH	4	1.108	(0.132)	1.385	(0.067)	0.181	(0.015)		
H052001	HH	8	0.581	(0.050)	-0.401	(0.146)	0.182	(0.046)		
H052401	HH	13	1.325	(0.144)	1.800	(0.084)	0.199	(0.011)		
H053501	HK	16	0.662	(0.026)	1.298	(0.036)	0.000	(0.000)	1.491	(0.046)
									-0.821	(0.078)
									-0.671	(0.130)
H053601	HK	17	0.644	(0.025)	1.180	(0.035)	0.000	(0.000)	1.785	(0.050)
									-0.441	(0.065)
									-1.343	(0.134)

Table D-18

IRT Parameters for the U.S. History Main Sample
Cultures Scale, Age 17/Grade 12

NAEP ID	Block	Item	A	S.E.	B	S.E.	C	S.E.	D	S.E.
H040601	HI	1	0.480	(0.064)	-0.033	(0.222)	0.270	(0.054)		
H040801	HI	3	0.891	(0.071)	-0.883	(0.120)	0.244	(0.045)		
H040802	HI	4	0.560	(0.047)	-1.536	(0.199)	0.234	(0.056)		
H041401	HI	12	0.653	(0.027)	0.266	(0.024)	0.000	(0.000)	0.740	(0.057)
									-0.172	(0.056)
									-0.568	(0.061)
H044201	HC	2	0.977	(0.111)	0.819	(0.062)	0.226	(0.022)		
H044301	HC	3	0.687	(0.027)	1.151	(0.029)	0.000	(0.000)	1.235	(0.048)
									0.328	(0.052)
									-1.563	(0.123)
H044601	HC	6	0.578	(0.060)	-0.706	(0.199)	0.269	(0.055)		
H045101	HC	13	0.759	(0.106)	0.681	(0.101)	0.308	(0.032)		
H045102	HC	14	0.976	(0.040)	0.768	(0.026)	0.000	(0.000)	0.714	(0.033)
									-0.714	(0.047)
H045301	HC	16	0.583	(0.024)	1.316	(0.048)	0.000	(0.000)	-1.521	(0.098)
									1.521	(0.112)
H045501	HD	1	0.582	(0.022)	0.985	(0.031)	0.000	(0.000)	1.409	(0.059)
									0.569	(0.056)
									-1.978	(0.124)
H046101	HD	7	0.243	(0.022)	2.995	(0.270)	0.000	(0.000)	-0.400	(0.136)
									0.400	(0.225)
H047401	HE	5	0.995	(0.053)	1.416	(0.043)	0.000	(0.000)	0.132	(0.038)
									-0.132	(0.066)
H047801	HE	9	1.276	(0.147)	0.997	(0.051)	0.232	(0.018)		
H047901	HE	10	1.218	(0.117)	0.407	(0.058)	0.295	(0.024)		
H048101	HE	12	1.490	(0.152)	0.786	(0.042)	0.254	(0.018)		
H049001	HF	5	0.430	(0.072)	0.490	(0.244)	0.296	(0.054)		
H049201	HF	8	1.055	(0.127)	1.214	(0.065)	0.154	(0.017)		
H049301	HF	9	0.982	(0.117)	0.818	(0.065)	0.256	(0.023)		
H049501	HF	11	1.083	(0.090)	-0.595	(0.096)	0.291	(0.038)		
H049502	HF	12	0.767	(0.103)	0.976	(0.083)	0.213	(0.027)		
H049503	HF	13	1.177	(0.059)	1.361	(0.033)	0.000	(0.000)	0.504	(0.030)
									-0.504	(0.063)
H050201	HG	3	0.493	(0.019)	0.617	(0.036)	0.000	(0.000)	1.726	(0.071)
									-0.280	(0.067)
									-1.446	(0.108)
H050202	HG	4	0.852	(0.094)	1.138	(0.068)	0.105	(0.018)		
H050203	HG	5	0.900	(0.224)	2.212	(0.263)	0.225	(0.017)		
H050801	HG	11	0.974	(0.102)	0.327	(0.076)	0.292	(0.028)		
H051101	HG	15	1.201	(0.049)	0.733	(0.022)	0.000	(0.000)	0.609	(0.028)
									-0.609	(0.039)
H051102	HG	16	1.063	(0.043)	0.234	(0.021)	0.000	(0.000)	0.294	(0.035)
									-0.294	(0.034)

Table D-18 (continued)

IRT Parameters for the U.S. History Main Sample
Cultures Scale, Age 17/Grade 12

NAEP ID	Block	Item	A	S.E.	B	S.E.	C	S.E.	D	S.E.
H051301	HH	1	0.792	(0.039)	1.094	(0.036)	0.000	(0.000)	0.462	(0.039)
									-0.462	(0.061)
H051701	HH	5	1.786	(0.149)	1.296	(0.048)	0.187	(0.011)		
H051801	HH	6	0.992	(0.098)	0.238	(0.076)	0.284	(0.029)		
H051901	HH	7	1.817	(0.142)	0.671	(0.029)	0.144	(0.014)		

Table D-19

IRT Parameters for the U.S. History Main Sample
Technology Scale, Age 17/Grade 12

NAEP ID	Block	Item	A	S.E.	B	S.E.	C	S.E.	D	S.E.
H041101	HI	8	0.993	(0.123)	0.989	(0.065)	0.224	(0.022)		
H041102	HI	9	0.656	(0.086)	0.419	(0.127)	0.296	(0.038)		
H041301	HI	11	0.478	(0.028)	1.485	(0.066)	0.000	(0.000)	0.785 -0.785	(0.060) (0.104)
H041501	HI	13	0.553	(0.053)	-1.758	(0.248)	0.302	(0.062)		
H042101	HJ	3	0.576	(0.069)	-0.453	(0.205)	0.327	(0.053)		
H042102	HJ	4	0.887	(0.098)	0.470	(0.083)	0.290	(0.030)		
H042701	HJ	11	0.506	(0.051)	-0.995	(0.212)	0.258	(0.053)		
H042702	HJ	12	0.893	(0.072)	-1.291	(0.131)	0.253	(0.047)		
H043001	HJ	16	0.570	(0.029)	0.412	(0.034)	0.000	(0.000)	0.472 -0.472	(0.056) (0.061)
H043101	HJ	17	0.328	(0.018)	0.215	(0.060)	0.000	(0.000)	1.358 -1.358	(0.101) (0.103)
H044101	HC	1	1.303	(0.184)	1.190	(0.065)	0.297	(0.018)		
H044801	HC	9	1.440	(0.138)	0.700	(0.042)	0.223	(0.019)		
H044902	HC	11	0.792	(0.093)	0.467	(0.095)	0.290	(0.032)		
H045601	HD	2	0.913	(0.092)	-0.031	(0.099)	0.344	(0.035)		
H045801	HD	4	1.529	(0.125)	0.261	(0.046)	0.276	(0.022)		
H046201	HD	8	1.659	(0.138)	0.359	(0.041)	0.275	(0.021)		
H047501	HE	6	0.394	(0.023)	0.514	(0.047)	0.000	(0.000)	0.224 -0.224	(0.078) (0.086)
H047601	HE	7	0.588	(0.026)	0.084	(0.034)	0.000	(0.000)	0.805 -0.805	(0.056) (0.054)
H048201	HE	13	1.405	(0.120)	-0.056	(0.062)	0.352	(0.028)		
H048202	HE	14	0.918	(0.037)	0.285	(0.025)	0.000	(0.000)	0.715 -0.715	(0.038) (0.039)
H048301	HE	15	0.785	(0.033)	0.232	(0.025)	0.000	(0.000)	0.109 -0.109	(0.045) (0.045)
H049901	HF	17	0.689	(0.097)	0.786	(0.108)	0.277	(0.033)		
H050001	HG	1	1.193	(0.090)	-0.308	(0.065)	0.243	(0.029)		
H050601	HG	9	1.171	(0.088)	-0.563	(0.073)	0.234	(0.031)		
H052101	HH	9	0.777	(0.069)	-0.628	(0.125)	0.263	(0.041)		
H052102	HH	10	1.519	(0.136)	0.495	(0.044)	0.275	(0.021)		
H052301	HH	12	0.536	(0.023)	1.328	(0.037)	0.000	(0.000)	1.557 0.191 -1.748	(0.058) (0.066) (0.159)
H052601	HH	15	0.495	(0.029)	1.649	(0.070)	0.000	(0.000)	0.807 -0.807	(0.059) (0.110)
H052901	HK	1	0.484	(0.020)	1.822	(0.037)	0.000	(0.000)	2.341 1.490 -3.830	(0.070) (0.064) (0.346)
H052902	HK	2	0.710	(0.026)	0.609	(0.033)	0.000	(0.000)	1.175 -1.175	(0.044) (0.059)
H053001	HK	3	1.330	(0.103)	-0.008	(0.054)	0.244	(0.026)		

Table D-19 (continued)

IRT Parameters for the U.S. History Main Sample
Technology Scale, Age 17/Grade 12

NAEP ID	Block	Item	A	S.E.	B	S.E.	C	S.E.	D	S.E.
H053002	HK	4	0.794	(0.099)	0.256	(0.117)	0.371	(0.036)		
H053003	HK	5	0.610	(0.037)	1.521	(0.065)	0.000	(0.000)	0.354	(0.052)
									-0.354	(0.091)
H053101	HK	6	0.580	(0.080)	1.218	(0.114)	0.162	(0.028)		
H053102	HK	7	0.710	(0.034)	1.037	(0.037)	0.000	(0.000)	0.690	(0.042)
									-0.690	(0.066)
H053103	HK	8	0.718	(0.035)	0.753	(0.032)	0.000	(0.000)	0.348	(0.044)
									-0.348	(0.055)
H053201	HK	9	0.600	(0.022)	-0.156	(0.038)	0.000	(0.000)	1.377	(0.063)
									-1.377	(0.054)
H053301	HK	10	1.465	(0.134)	-0.841	(0.089)	0.413	(0.036)		
H053302	HK	11	1.630	(0.178)	0.558	(0.050)	0.398	(0.021)		
H053303	HK	12	0.898	(0.097)	0.260	(0.090)	0.333	(0.031)		
H053304	HK	13	0.807	(0.037)	0.683	(0.028)	0.000	(0.000)	0.288	(0.040)
									-0.288	(0.049)
H053401	HK	14	0.811	(0.143)	1.333	(0.114)	0.318	(0.026)		
H053402	HK	15	0.456	(0.056)	0.182	(0.178)	0.223	(0.044)		
H053701	HK	18	0.610	(0.039)	1.822	(0.071)	0.000	(0.000)	0.857	(0.054)
									-0.857	(0.128)

Table D-20

IRT Parameters for the U.S. History Main Sample
World Role Scale, Age 17/Grade 12

NAEP ID	Block	Item	A	S.E.	B	S.E.	C	S.E.	D	S.E.
H041001	HI	7	1.382	(0.147)	1.018	(0.048)	0.247	(0.017)		
H041201	HI	10	0.847	(0.078)	0.331	(0.081)	0.200	(0.029)		
H041601	HI	14	1.932	(0.157)	0.797	(0.034)	0.251	(0.015)		
H041701	HI	15	2.117	(0.123)	1.206	(0.032)	0.137	(0.010)		
H041901	HJ	1	0.979	(0.169)	1.474	(0.099)	0.345	(0.020)		
H042201	HJ	5	0.927	(0.065)	2.185	(0.090)	0.000	(0.000)	0.085 -0.085	(0.056) (0.131)
H042301	HJ	6	1.854	(0.140)	1.325	(0.045)	0.245	(0.012)		
H042401	HJ	7	1.497	(0.176)	1.849	(0.098)	0.272	(0.012)		
H042901	HJ	14	1.811	(0.127)	-0.574	(0.048)	0.205	(0.027)		
H042902	HJ	15	0.958	(0.036)	0.126	(0.022)	0.000	(0.000)	0.173 -0.173	(0.040) (0.038)
H044701	HC	7	1.110	(0.100)	0.664	(0.052)	0.18[illegible]	(0.020)		
H044702	HC	8	1.058	(0.059)	1.659	(0.047)	0.0[illegible]	(0.000)	0.249 -0.249	(0.039) (0.077)
H045001	HC	12	0.803	(0.091)	0.542	(0.093)	0.253	(0.031)		
H045201	HC	15	1.758	(0.135)	1.253	(0.043)	0.193	(0.012)		
H045401	HC	17	1.787	(0.155)	1.565	(0.061)	0.231	(0.011)		
H046301	HD	9	0.571	(0.030)	1.337	(0.055)	0.000	(0.000)	0.392 -0.392	(0.055) (0.086)
H046401	HD	10	1.992	(0.140)	1.370	(0.043)	0.195	(0.010)		
H046402	HD	11	1.050	(0.112)	0.797	(0.061)	0.238	(0.022)		
H046601	HD	13	1.158	(0.114)	0.844	(0.051)	0.193	(0.018)		
H046701	HD	14	0.758	(0.104)	1.111	(0.090)	0.225	(0.026)		
H046702	HD	15	1.090	(0.097)	0.798	(0.049)	0.131	(0.017)		
H046801	HD	16	1.172	(0.103)	0.569	(0.051)	0.184	(0.020)		
H047001	HE	1	0.445	(0.053)	-0.031	(0.225)	0.238	(0.054)		
H047701	HE	8	1.073	(0.130)	0.812	(0.071)	0.355	(0.023)		
H048001	HE	11	1.363	(0.132)	1.374	(0.055)	0.178	(0.012)		
H048401	HE	16	0.990	(0.111)	0.543	(0.080)	0.324	(0.027)		
H048501	HE	17	2.062	(0.144)	1.227	(0.039)	0.247	(0.012)		
H049601	HF	14	0.838	(0.032)	1.499	(0.031)	0.000	(0.000)	1.530 -0.428 -1.102	(0.038) (0.062) (0.156)
H049701	HF	15	1.153	(0.043)	1.065	(0.023)	0.000	(0.000)	-0.704 0.704	(0.056) (0.062)
H049801	HF	16	1.067	(0.129)	1.302	(0.065)	0.168	(0.017)		
H050301	HG	6	0.683	(0.081)	0.971	(0.090)	0.172	(0.028)		
H050701	HG	10	1.040	(0.134)	1.016	(0.071)	0.316	(0.022)		
H050901	HG	12	1.518	(0.135)	0.746	(0.042)	0.228	(0.017)		
H051001	HG	13	1.992	(0.129)	1.484	(0.043)	0.143	(0.009)		
H051002	HG	14	1.867	(0.104)	1.541	(0.027)	0.000	(0.000)	0.296 -0.296	(0.026) (0.050)
H051201	HG	17	1.068	(0.105)	0.853	(0.056)	0.192	(0.020)		

Table D-20 (continued)

IRT Parameters for the U.S. History Main Sample
World Role Scale, Age 17/Grade 12

NAEP ID	Block	Item	A	S.E.	B	S.E.	C	S.E.	D	S.E.
H052201	HH	11	0.636	(0.068)	0.122	(0.139)	0.232	(0.043)		
H052501	HH	14	0.946	(0.040)	1.301	(0.034)	0.000	(0.000)	-0.232	(0.046)
									0.232	(0.061)
H052701	HH	16	0.488	(0.020)	0.447	(0.037)	0.000	(0.000)	-0.271	(0.069)
									0.271	(0.074)
H052801	HH	17	1.028	(0.097)	0.262	(0.076)	0.279	(0.029)		

Table D-21

IRT Parameters for the Geography Main Samples
Space and Place Scale, Age 9/Grade 4

NAEP ID	Block	Item	A	S.E.	B	S.E.	C	S.E.	D	S.E.
G007001	GC	1	0.831	(0.028)	0.379	(0.022)	0.000	(0.000)	1.240 0.199 -1.439	(0.046) (0.039) (0.057)
G007002	GC	2	0.893	(0.097)	1.137	(0.064)	0.162	(0.018)		
G007003	GC	3	1.707	(0.151)	0.682	(0.038)	0.287	(0.017)		
G007004	GC	4	1.254	(0.047)	0.605	(0.018)	0.000	(0.000)	0.167 -0.167	(0.029) (0.032)
G007501	GC	10	1.186	(0.114)	0.528	(0.059)	0.307	(0.022)		
G007601	GC	11	0.570	(0.023)	-0.485	(0.036)	0.000	(0.000)	0.146 -0.146	(0.066) (0.055)
G007701	GC	12	0.455	(0.019)	-0.618	(0.036)	0.000	(0.000)	0.626 -0.117 -0.508	(0.099) (0.084) (0.071)
G008001	GD	1	0.676	(0.022)	0.028	(0.026)	0.000	(0.000)	-0.426 0.426	(0.056) (0.054)
G008002	GD	2	1.784	(0.133)	0.946	(0.032)	0.177	(0.012)		
G008201	GD	4	0.413	(0.016)	1.102	(0.041)	0.000	(0.000)	-0.404 0.935 -0.532	(0.085) (0.100) (0.115)
G008301	GD	5	1.125	(0.186)	1.165	(0.083)	0.474	(0.020)		
G008401	GD	6	0.650	(0.079)	0.441	(0.123)	0.283	(0.036)		
G008501	GD	7	1.974	(0.136)	0.959	(0.031)	0.216	(0.012)		
G008502	GD	8	1.132	(0.122)	0.916	(0.055)	0.260	(0.019)		
G008601	GD	10	0.712	(0.068)	-0.466	(0.147)	0.300	(0.045)		
G008701	GD	11	0.581	(0.026)	0.461	(0.034)	0.000	(0.000)	0.641 -0.641	(0.053) (0.060)
G008902	GD	14	1.265	(0.107)	-0.188	(0.072)	0.335	(0.029)		
G009101	GE	1	0.831	(0.071)	-0.195	(0.101)	0.270	(0.034)		
G009102	GE	2	1.097	(0.084)	0.475	(0.048)	0.165	(0.019)		
G009103	GE	3	1.225	(0.101)	0.119	(0.061)	0.313	(0.024)		
G009401	GE	6	1.394	(0.119)	0.955	(0.038)	0.150	(0.013)		

Table D-21 (continued)

IRT Parameters for the Geography Main Samples
Space and Place Scale, Age 9/Grade 4

NAEP ID	Block	Item	A	S.E.	B	S.E.	C	S.E.	D	S.E.
G009601	GE	11	0.574	(0.019)	0.018	(0.024)	0.000	(0.000)	-0.122	(0.070)
									0.175	(0.073)
									-0.053	(0.062)
G010201	GF	2	0.629	(0.056)	-0.888	(0.180)	0.284	(0.053)		
G011201	GG	2	0.827	(0.073)	-0.202	(0.107)	0.284	(0.036)		
G011301	GG	3	1.383	(0.136)	0.975	(0.044)	0.232	(0.016)		
G011701	GG	9	0.734	(0.180)	2.104	(0.232)	0.323	(0.022)		
G011901	GG	11	1.558	(0.148)	1.243	(0.052)	0.221	(0.014)		
G012101	GG	13	1.692	(0.156)	1.009	(0.040)	0.109	(0.013)		
G012102	GG	14	1.605	(0.187)	1.506	(0.078)	0.239	(0.014)		
G012201	GH	1	0.748	(0.033)	-0.657	(0.044)	0.000	(0.000)		
G012202	GH	2	1.295	(0.111)	0.596	(0.046)	0.247	(0.018)		
G012203	GH	3	1.456	(0.165)	0.568	(0.061)	0.498	(0.019)		
G012301	GH	4	1.369	(0.125)	0.523	(0.050)	0.327	(0.019)		
G012401	GH	6	0.563	(0.066)	0.285	(0.146)	0.257	(0.040)		
G012601	GH	10	1.445	(0.126)	0.776	(0.040)	0.225	(0.016)		
G013001	GH	15	0.702	(0.024)	1.179	(0.028)	0.000	(0.000)	-0.720	(0.067)
									0.851	(0.081)
									-0.130	(0.080)
G013101	GH	16	0.830	(0.115)	1.457	(0.089)	0.177	(0.019)		

Table D-22

IRT Parameters for the Geography Main Samples
Environment and Society Scale, Age 9/Grade 4

NAEP ID	Block	Item	A	S.E.	B	S.E.	C	S.E.	D	S.E.
G007101	GC	5	0.819	(0.098)	0.839	(0.073)	0.245	(0.024)		
G007201	GC	6	0.558	(0.026)	0.905	(0.040)	0.000	(0.000)	0.763 -0.763	(0.052) (0.070)
G007301	GC	7	0.239	(0.051)	2.608	(0.504)	0.331	(0.036)		
G007401	GC	8	1.055	(0.077)	-0.115	(0.064)	0.197	(0.025)		
G007402	GC	9	1.057	(0.084)	-0.494	(0.089)	0.299	(0.032)		
G007801	GC	13	1.015	(0.090)	-0.373	(0.099)	0.332	(0.034)		
G008101	GD	3	1.091	(0.092)	0.282	(0.059)	0.250	(0.023)		
G008801	GD	12	1.495	(0.143)	1.015	(0.044)	0.281	(0.016)		
G008901	GD	13	0.696	(0.080)	-0.166	(0.153)	0.356	(0.041)		
G009001	GD	15	0.470	(0.020)	-0.428	(0.034)	0.000	(0.000)	1.597 -0.499 -1.097	(0.097) (0.072) (0.074)
G009201	GE	4	0.628	(0.029)	0.981	(0.037)	0.000	(0.000)	0.595 -0.595	(0.046) (0.063)
G009801	GE	13	0.839	(0.066)	-1.267	(0.143)	0.287	(0.048)		
G009901	GE	14	0.529	(0.064)	0.043	(0.173)	0.261	(0.044)		
G010101	GF	1	1.304	(0.097)	0.108	(0.052)	0.244	(0.022)		
G010401	GF	4	1.316	(0.128)	0.641	(0.049)	0.312	(0.020)		
G010402	GF	5	1.313	(0.107)	0.328	(0.049)	0.270	(0.021)		
G010601	GF	8	0.833	(0.099)	0.841	(0.071)	0.240	(0.024)		
G010602	GF	9	0.301	(0.028)	1.649	(0.159)	0.000	(0.000)		
G010603	GF	10	0.410	(0.033)	1.477	(0.114)	0.000	(0.000)		
G010701	GF	11	1.174	(0.054)	1.126	(0.025)	0.000	(0.000)	0.137 -0.137	(0.031) (0.043)
G010901	GF	14	0.665	(0.027)	0.890	(0.035)	0.000	(0.000)	0.970 -0.970	(0.046) (0.064)
G011801	GG	10	0.792	(0.038)	0.890	(0.032)	0.000	(0.000)	0.535 -0.535	(0.042) (0.056)
G012302	GH	5	1.145	(0.150)	1.607	(0.091)	0.225	(0.014)		
G012801	GH	12	0.580	(0.057)	-0.364	(0.161)	0.251	(0.043)		
G013201	GH	17	0.994	(0.059)	1.484	(0.047)	0.000	(0.000)	0.054 -0.054	(0.043) (0.071)

Table D-23

IRT Parameters for the Geography Main Samples
Spatial Dynamics and Connections Scale, Age 9/Grade 4

NAEP ID	Block	Item	A S.E.	B S.E.	C S.E.	D S.E.
G007901	GC	14	0.546 (0.026)	1.317 (0.052)	0.000 (0.000)	-0.649 (0.077) 0.649 (0.095)
G008503	GD	9	0.890 (0.035)	0.388 (0.023)	0.000 (0.000)	0.362 (0.038) -0.362 (0.039)
G009301	GE	5	1.149 (0.098)	0.534 (0.053)	0.250 (0.021)	
G009402	GE	7	0.727 (0.025)	-0.584 (0.032)	0.000 (0.000)	1.476 (0.062) -1.476 (0.040)
G009403	GE	8	0.754 (0.024)	0.370 (0.031)	0.000 (0.000)	1.347 (0.044) -1.347 (0.050)
G009501	GE	9	0.760 (0.092)	0.469 (0.109)	0.372 (0.031)	
G009502	GE	10	1.915 (0.139)	0.893 (0.033)	0.269 (0.014)	
G010001	GE	15	1.704 (0.148)	0.638 (0.040)	0.268 (0.018)	
G010301	GF	3	0.568 (0.026)	1.048 (0.041)	0.000 (0.000)	0.822 (0.050) -0.822 (0.073)
G010501	GF	6	1.456 (0.129)	1.079 (0.041)	0.193 (0.014)	
G010502	GF	7	0.521 (0.026)	1.879 (0.080)	0.000 (0.000)	-0.741 (0.080) 0.741 (0.115)
G010801	GF	12	0.493 (0.059)	0.300 (0.162)	0.234 (0.041)	
G010802	GF	13	0.854 (0.086)	0.395 (0.084)	0.281 (0.029)	
G011001	GF	15	0.988 (0.078)	-0.138 (0.076)	0.234 (0.029)	
G011101	GG	1	1.133 (0.103)	0.615 (0.055)	0.260 (0.021)	
G011302	GG	4	1.359 (0.128)	1.221 (0.047)	0.177 (0.013)	
G011303	GG	5	1.023 (0.164)	1.682 (0.106)	0.225 (0.016)	
G011401	GG	6	0.909 (0.047)	0.779 (0.037)	0.000 (0.000)	
G011501	GG	7	0.731 (0.030)	1.176 (0.030)	0.000 (0.000)	1.041 (0.043) -0.249 (0.057) -0.791 (0.099)
G011601	GG	8	0.933 (0.041)	0.984 (0.029)	0.000 (0.000)	0.596 (0.035) -0.596 (0.051)
G012001	GG	12	1.263 (0.079)	1.503 (0.043)	0.000 (0.000)	0.262 (0.038) -0.262 (0.072)
G012501	GH	7	1.566 (0.295)	2.315 (0.213)	0.233 (0.010)	
G012502	GH	8	1.334 (0.131)	1.437 (0.061)	0.239 (0.013)	
G012503	GH	9	0.908 (0.037)	1.054 (0.028)	0.000 (0.000)	0.671 (0.033) -0.671 (0.052)
G012701	GH	11	1.175 (0.213)	2.018 (0.171)	0.364 (0.013)	
G012901	GH	13	1.558 (0.122)	1.120 (0.039)	0.195 (0.013)	
G012902	GH	14	0.727 (0.039)	0.727 (0.043)	0.000 (0.000)	

Table D-24

IRT Parameters for the Geography Main Samples
Space and Place Scale, Age 13/Grade 8

NAEP ID	Block	Item	A	S.E.	B	S.E.	C	S.E.	D	S.E.
G011201	GG	2	0.657	(0.052)	-2.015	(0.172)	0.201	(0.056)		
G011301	GG	3	0.846	(0.056)	-0.340	(0.075)	0.129	(0.030)		
G011701	GG	9	0.540	(0.052)	-0.124	(0.153)	0.187	(0.045)		
G011901	GG	11	1.463	(0.100)	-0.131	(0.046)	0.179	(0.023)		
G012101	GG	13	1.178	(0.076)	-0.217	(0.053)	0.124	(0.024)		
G012102	GG	14	0.971	(0.091)	0.448	(0.070)	0.225	(0.027)		
G012201	GH	1	0.417	(0.016)	-1.854	(0.072)	0.000	(0.000)	-1.963	(0.150)
									1.963	(0.124)
G012202	GH	2	1.032	(0.082)	-0.497	(0.092)	0.264	(0.039)		
G012203	GH	3	0.790	(0.061)	-1.094	(0.135)	0.227	(0.053)		
G012301	GH	4	0.759	(0.052)	-1.400	(0.122)	0.178	(0.048)		
G012401	GH	6	0.495	(0.049)	-0.460	(0.198)	0.215	(0.054)		
G012601	GH	10	0.976	(0.062)	-0.575	(0.071)	0.138	(0.032)		
G013001	GH	15	0.549	(0.016)	-0.133	(0.024)	0.000	(0.000)	-0.906	(0.086)
									0.673	(0.091)
									0.234	(0.065)
G013101	GH	16	1.133	(0.097)	0.117	(0.067)	0.282	(0.028)		
G013301	GC	1	1.001	(0.071)	-1.380	(0.105)	0.201	(0.049)		
G013302	GC	2	0.532	(0.044)	-1.278	(0.178)	0.192	(0.053)		
G013401	GC	3	0.600	(0.045)	-1.084	(0.143)	0.174	(0.047)		
G013601	GC	7	0.754	(0.054)	-1.230	(0.123)	0.180	(0.048)		
G013602	GC	8	0.675	(0.053)	-0.461	(0.118)	0.168	(0.042)		
G013701	GC	9	0.942	(0.072)	-0.033	(0.073)	0.184	(0.030)		
G014101	GC	13	0.840	(0.092)	1.361	(0.075)	0.122	(0.018)		
G014401	GC	16	0.851	(0.030)	1.182	(0.032)	0.000	(0.000)	1.468	(0.036)
									-1.468	(0.081)
G014501	GC	17	0.703	(0.123)	1.742	(0.137)	0.264	(0.024)		
G014601	GD	1	0.953	(0.074)	-0.969	(0.115)	0.247	(0.050)		
G015101	GD	6	0.858	(0.068)	0.087	(0.078)	0.171	(0.031)		
G016101	GD	16	1.670	(0.129)	0.608	(0.035)	0.187	(0.016)		
G016102	GD	17	0.850	(0.103)	0.982	(0.084)	0.272	(0.026)		
G016201	GE	1	0.926	(0.042)	-0.293	(0.033)	0.000	(0.000)		
G016202	GE	2	1.014	(0.065)	-0.688	(0.074)	0.152	(0.034)		
G016203	GE	3	0.667	(0.062)	-0.865	(0.183)	0.272	(0.061)		
G016204	GE	4	1.007	(0.072)	-0.876	(0.094)	0.210	(0.044)		
G016301	GE	5	1.153	(0.080)	0.187	(0.048)	0.131	(0.021)		
G016501	GE	8	0.709	(0.053)	-0.170	(0.092)	0.138	(0.033)		
G016901	GE	13	0.610	(0.074)	0.802	(0.116)	0.201	(0.035)		
G017501	GF	2	0.618	(0.087)	0.571	(0.156)	0.323	(0.043)		
G017601	GF	3	0.794	(0.055)	-1.336	(0.112)	0.168	(0.045)		
G017602	GF	4	0.933	(0.072)	-0.344	(0.088)	0.197	(0.037)		

Table D-24 (continued)

IRT Parameters for the Geography Main Samples
Space and Place Scale, Age 13/Grade 8

NAEP ID	Block	Item	A	S.E.	B	S.E.	C	S.E.	D	S.E.
G018001	GF	8	1.109	(0.099)	0.563	(0.057)	0.222	(0.022)		
G018301	GF	11	0.600	(0.023)	0.540	(0.027)	0.000	(0.000)	0.401	(0.059)
									0.344	(0.063)
									-0.746	(0.073)
G018601	GF	14	1.204	(0.159)	2.158	(0.125)	0.176	(0.010)		
G018701	GF	15	0.858	(0.074)	0.446	(0.069)	0.147	(0.026)		
G018801	GF	16	0.528	(0.022)	0.012	(0.035)	0.000	(0.000)	-0.390	(0.073)
									0.390	(0.072)
G018901	GI	1	1.051	(0.068)	-0.151	(0.057)	0.131	(0.025)		
G019001	GI	2	0.773	(0.059)	-2.028	(0.153)	0.205	(0.057)		
G019002	GI	3	0.461	(0.022)	1.687	(0.073)	0.000	(0.000)	-0.603	(0.079)
								0.603	(0.111)	
G019301	GI	9	1.018	(0.074)	0.217	(0.056)	0.138	(0.023)		
G019302	GI	10	0.547	(0.024)	-0.257	(0.036)	0.000	(0.000)	0.886	(0.061)
									-0.886	(0.054)
G019401	GI	11	1.132	(0.088)	0.949	(0.043)	0.086	(0.013)		
G019402	GI	12	0.556	(0.024)	0.389	(0.033)	0.000	(0.000)	0.190	(0.055)
									-0.190	(0.061)
G019403	GI	13	0.394	(0.052)	0.526	(0.220)	0.200	(0.050)		
G019501	GI	14	0.476	(0.043)	-0.386	(0.168)	0.173	(0.046)		
G019601	GI	15	0.789	(0.074)	0.334	(0.089)	0.209	(0.032)		
G019701	GI	16	0.426	(0.055)	0.836	(0.176)	0.179	(0.043)		
G019702	GI	17	0.652	(0.075)	0.083	(0.154)	0.288	(0.047)		
G019703	GI	18	0.773	(0.093)	0.863	(0.091)	0.260	(0.029)		
G019801	GI	19	0.813	(0.063)	-0.456	(0.103)	0.198	(0.040)		
G020101	GI	22	0.655	(0.075)	0.688	(0.107)	0.214	(0.034)		
G020401	GI	26	0.244	(0.032)	-0.056	(0.305)	0.193	(0.047)		

BEST COPY AVAILABLE

Table D-25

IRT Parameters for the Geography Main Samples
Environment and Society Scale, Age 13/Grade 8

NAEP ID	Block	Item	A	S.E.	B	S.E.	C	S.E.	D	S.E.
G011801	GG	10	0.635	(0.027)	-0.677	(0.037)	0.000	(0.000)	0.389 -0.389	(0.064) (0.048)
G012302	GH	5	1.046	(0.105)	0.628	(0.056)	0.237	(0.022)		
G012801	GH	12	0.489	(0.043)	-1.765	(0.241)	0.268	(0.060)		
G013201	GH	17	0.673	(0.028)	0.359	(0.027)	0.000	(0.000)	0.023 -0.023	(0.051) (0.052)
G013801	GC	10	2.645	(0.160)	1.254	(0.035)	0.253	(0.011)		
G014001	GC	12	0.931	(0.046)	1.349	(0.040)	0.000	(0.000)	-0.115 0.115	(0.043) (0.064)
G014301	GC	15	0.754	(0.035)	0.875	(0.032)	0.000	(0.000)	0.303 -0.303	(0.042) (0.055)
G014701	GD	2	1.309	(0.105)	-0.226	(0.068)	0.320	(0.028)		
G014801	GD	3	1.337	(0.149)	1.253	(0.058)	0.235	(0.015)		
G014901	GD	4	0.876	(0.040)	0.959	(0.029)	0.000	(0.000)	0.367 -0.367	(0.036) (0.050)
G015201	GD	7	0.715	(0.046)	1.170	(0.062)	0.000	(0.000)		
G015301	GD	8	0.706	(0.037)	1.276	(0.046)	0.000	(0.000)	0.156 -0.156	(0.046) (0.069)
G015501	GD	10	1.368	(0.112)	0.380	(0.045)	0.237	(0.020)		
G015601	GD	11	1.280	(0.168)	1.362	(0.076)	0.310	(0.016)		
G015801	GD	13	2.101	(0.172)	0.459	(0.032)	0.285	(0.017)		
G016401	GE	7	0.695	(0.027)	0.307	(0.026)	0.000	(0.000)	-0.021 0.021	(0.049) (0.049)
G016701	GE	11	0.855	(0.036)	0.547	(0.025)	0.000	(0.000)	0.459 -0.459	(0.037) (0.043)
G017001	GE	14	2.220	(0.149)	0.741	(0.025)	0.190	(0.013)		
G017002	GE	15	1.460	(0.130)	0.608	(0.040)	0.239	(0.018)		
G017401	GF	1	0.708	(0.068)	-0.477	(0.147)	0.278	(0.045)		
G018201	GF	10	1.672	(0.161)	1.231	(0.053)	0.325	(0.014)		
G018501	GF	13	0.572	(0.025)	1.842	(0.037)	0.000	(0.000)	1.130 1.030 -2.160	(0.056) (0.067) (0.236)
G019003	GI	4	0.603	(0.026)	-0.748	(0.038)	0.000	(0.000)	0.946 -0.946	(0.069) (0.046)
G019101	GI	5	1.511	(0.132)	0.440	(0.044)	0.305	(0.020)		
G019102	GI	6	0.927	(0.052)	1.569	(0.052)	0.000	(0.000)	0.071 -0.071	(0.041) (0.074)
G019201	GI	7	0.686	(0.087)	0.321	(0.127)	0.309	(0.038)		
G019202	GI	8	0.628	(0.033)	1.369	(0.049)	0.000	(0.000)	0.576 -0.576	(0.046) (0.079)
G019901	GI	20	0.337	(0.011)	0.288	(0.047)	0.000	(0.000)	1.999 1.016 -3.015	(0.119) (0.087) (0.132)

Table D-25 (continued)

IRT Parameters for the Geography Main Samples
Environment and Society Scale, Age 13/Grade 8

NAEP ID	Block	Item	A	S.E.	B	S.E.	C	S.E.	D	S.E.
G020001	GI	21	1.231	(0.084)	1.999	(0.063)	0.000	(0.000)	0.438	(0.036)
									-0.438	(0.130)
G020201	GI	23	0.690	(0.024)	1.401	(0.037)	0.000	(0.000)	1.612	(0.041)
									-1.612	(0.101)
G020301	GI	24	3.865	(0.164)	0.902	(0.018)	0.214	(0.011)		
G020302	GI	25	1.036	(0.046)	1.068	(0.027)	0.000	(0.000)	0.050	(0.034)
									-0.050	(0.047)

Table D-26

IRT Parameters for the Geography Main Samples
Spatial Dynamics and Connections Scale, Age 13/Grade 8

NAEP ID	Block	Item	A	S.E.	B	S.E.	C	S.E.	D	S.E.
G011101	GG	1	0.890	(0.082)	-0.440	(0.121)	0.355	(0.042)		
G011302	GG	4	1.395	(0.103)	0.206	(0.046)	0.220	(0.021)		
G011303	GG	5	1.198	(0.126)	0.736	(0.058)	0.302	(0.021)		
G011401	GG	6	0.884	(0.041)	-0.738	(0.042)	0.000	(0.000)		
G011501	GG	7	0.532	(0.022)	0.096	(0.029)	0.000	(0.000)	1.097	(0.071)
									0.017	(0.062)
									-1.113	(0.071)
G011601	GG	8	0.788	(0.032)	-0.235	(0.027)	0.000	(0.000)	0.672	(0.047)
									-0.672	(0.039)
G012001	GG	12	0.989	(0.038)	0.081	(0.022)	0.000	(0.000)	0.226	(0.038)
									-0.226	(0.036)
G012501	GH	7	1.309	(0.134)	0.963	(0.047)	0.226	(0.017)		
G012502	GH	8	1.279	(0.122)	0.425	(0.060)	0.355	(0.023)		
G012503	GH	9	0.854	(0.043)	-1.413	(0.060)	0.000	(0.000)		
G012701	GH	11	0.992	(0.111)	0.546	(0.079)	0.354	(0.027)		
G012901	GH	13	1.335	(0.098)	-0.063	(0.054)	0.248	(0.025)		
G012902	GH	14	0.735	(0.036)	-0.486	(0.043)	0.000	(0.000)		
G013402	GC	4	0.796	(0.041)	1.658	(0.049)	0.000	(0.000)	0.653	(0.040)
									-0.653	(0.090)
G013501	GC	5	1.741	(0.180)	0.769	(0.045)	0.376	(0.018)		
G013502	GC	6	1.695	(0.152)	0.846	(0.037)	0.222	(0.015)		
G013901	GC	11	1.087	(0.087)	0.253	(0.059)	0.209	(0.025)		
G014201	GC	14	0.685	(0.029)	0.984	(0.028)	0.000	(0.000)	0.426	(0.048)
									-0.004	(0.062)
									-0.422	(0.081)
G015001	GD	5	1.155	(0.110)	0.655	(0.056)	0.264	(0.022)		
G015401	GD	9	1.128	(0.090)	-0.408	(0.084)	0.311	(0.035)		
G015701	GD	12	1.369	(0.138)	1.159	(0.049)	0.253	(0.015)		
G015901	GD	14	0.511	(0.021)	-0.095	(0.034)	0.000	(0.000)	-0.109	(0.067)
									0.109	(0.063)
G016001	GD	15	1.137	(0.053)	1.593	(0.031)	0.000	(0.000)	0.169	(0.038)
									0.217	(0.058)
									-0.386	(0.098)
G016302	GE	6	1.262	(0.068)	1.596	(0.039)	0.000	(0.000)	0.040	(0.037)
									-0.040	(0.064)
G016502	GE	9	0.862	(0.048)	1.797	(0.056)	0.000	(0.000)	0.441	(0.041)
									-0.441	(0.093)
G016601	GE	10	1.025	(0.122)	1.039	(0.063)	0.251	(0.021)		
G016801	GE	12	0.506	(0.102)	1.615	(0.175)	0.274	(0.036)		
G017101	GE	16	0.905	(0.037)	1.136	(0.025)	0.000	(0.000)	0.528	(0.036)
									-0.525	(0.062)
									-0.003	(0.078)
G017301	GE	18	1.187	(0.127)	0.744	(0.060)	0.313	(0.022)		

Table D-26 (continued)

IRT Parameters for the Geography Main Samples
Spatial Dynamics and Connections Scale, Age 13/Grade 8

NAEP ID	Block	Item	A	S.E.	B	S.E.	C	S.E.	D	S.E.
G017701	GF	5	0.494	(0.016)	0.508	(0.034)	0.000	(0.000)	-1.380	(0.087)
									1.380	(0.091)
G017801	GF	6	0.948	(0.095)	0.045	(0.098)	0.361	(0.033)		
G017901	GF	7	0.554	(0.074)	0.610	(0.147)	0.257	(0.040)		
G018101	GF	9	0.877	(0.043)	1.322	(0.040)	0.000	(0.000)	0.253	(0.040)
									-0.253	(0.065)
G018401	GF	12	0.478	(0.129)	2.078	(0.301)	0.401	(0.034)		

Table D-27

IRT Parameters for the Geography Main Samples
Space and Place Scale, Age 17/Grade 12

NAEP ID	Block	Item	A	S.E.	B	S.E.	C	S.E.	D	S.E.
G016201	GE	1	0.681	(0.023)	-0.918	(0.034)	0.000	(0.000)	-0.793	(0.077)
									0.793	(0.067)
016202	GE	2	1.046	(0.078)	-1.193	(0.104)	0.255	(0.048)		
G016203	GE	3	0.718	(0.061)	-1.282	(0.169)	0.288	(0.058)		
G016204	GE	4	0.841	(0.067)	-1.677	(0.149)	0.273	(0.058)		
G016301	GE	5	1.242	(0.089)	-0.483	(0.069)	0.236	(0.033)		
G016501	GE	8	0.539	(0.049)	-0.759	(0.183)	0.243	(0.052)		
G016901	GE	13	0.942	(0.110)	0.516	(0.089)	0.342	(0.029)		
G017501	GF	2	0.721	(0.087)	-0.016	(0.160)	0.377	(0.045)		
G017601	GF	3	0.858	(0.067)	-1.774	(0.141)	0.256	(0.056)		
G017602	GF	4	0.856	(0.064)	-0.626	(0.102)	0.210	(0.040)		
G018001	GF	8	1.347	(0.103)	-0.084	(0.059)	0.264	(0.026)		
G018301	GF	11	0.518	(0.020)	-0.013	(0.029)	0.000	(0.000)	0.173	(0.081)
									0.681	(0.076)
									-0.854	(0.069)
G018601	GF	14	0.998	(0.166)	1.883	(0.129)	0.143	(0.014)		
G018701	GF	15	0.915	(0.072)	-0.169	(0.085)	0.210	(0.033)		
G018801	GF	16	0.526	(0.021)	-0.831	(0.045)	0.000	(0.000)	-0.545	(0.088)
									0.545	(0.074)
G020601	GC	2	0.842	(0.080)	-0.065	(0.110)	0.288	(0.038)		
G020602	GC	3	0.647	(0.127)	1.671	(0.147)	0.250	(0.028)		
G020801	GC	5	0.883	(0.074)	-0.132	(0.093)	0.228	(0.035)		
G021101	GC	9	0.324	(0.042)	-0.644	(0.335)	0.312	(0.057)		
G021401	GC	12	0.693	(0.021)	-0.184	(0.022)	0.000	(0.000)	-0.282	(0.067)
									0.189	(0.072)
									0.092	(0.056)
G021601	GC	14	0.957	(0.051)	1.832	(0.046)	0.000	(0.000)	0.308	(0.045)
									0.199	(0.078)
									-0.508	(0.153)
G021901	GD	2	0.856	(0.071)	-0.485	(0.114)	0.261	(0.043)		
G022201	GD	5	0.678	(0.026)	0.798	(0.029)	0.000	(0.000)	1.267	(0.048)
									-0.694	(0.061)
									-0.573	(0.086)
G023201	GG	1	0.831	(0.061)	-0.959	(0.116)	0.232	(0.046)		
G022401	GD	7	0.799	(0.069)	0.008	(0.097)	0.208	(0.035)		
G022402	GD	8	0.879	(0.094)	0.427	(0.090)	0.289	(0.030)		

Table D-27 (continued)

IRT Parameters for the Geography Main Samples
Space and Place Scale, Age 17/Grade 12

NAEP ID	Block	Item	A	S.E.	B	S.E.	C	S.E.	D	S.E.
G023301	GG	2	1.232	(0.100)	-1.079	(0.099)	0.319	(0.048)		
G023401	GG	3	1.294	(0.123)	-0.251	(0.090)	0.444	(0.034)		
G023701	GG	6	0.907	(0.065)	-0.652	(0.096)	0.212	(0.040)		
G024001	GG	9	0.926	(0.066)	-0.378	(0.084)	0.189	(0.035)		
G024301	GG	12	0.680	(0.031)	0.565	(0.032)	0.000	(0.000)	0.514 -0.514	(0.049) (0.056)
G024401	GG	13	0.382	(0.072)	0.730	(0.317)	0.332	(0.060)		
G024501	GG	14	1.212	(0.102)	-0.318	(0.083)	0.283	(0.036)		
G024601	GG	15	0.650	(0.037)	0.471	(0.043)	0.000	(0.000)	0.898 -0.898	(0.067) (0.071)
G024701	GH	1	1.625	(0.156)	0.891	(0.041)	0.270	(0.016)		
G024901	GH	3	1.981	(0.155)	1.730	(0.061)	0.136	(0.009)		
G024902	GH	4	1.159	(0.173)	1.425	(0.084)	0.336	(0.017)		
G024903	GH	5	1.017	(0.112)	0.164	(0.100)	0.399	(0.033)		
G025401	GH	11	3.251	(0.214)	0.542	(0.021)	0.177	(0.012)		
G025402	GH	12	3.755	(0.465)	0.386	(0.038)	0.623	(0.015)		
G025901	GI	1	1.299	(0.136)	0.927	(0.047)	0.221	(0.017)		
G026001	GI	2	1.731	(0.131)	1.091	(0.037)	0.201	(0.012)		
G026002	GI	3	1.085	(0.178)	1.402	(0.088)	0.364	(0.019)		
G026201	GI	5	1.483	(0.122)	0.159	(0.054)	0.307	(0.023)		
G026202	GI	6	0.830	(0.078)	0.230	(0.091)	0.231	(0.032)		
G026203	GI	7	1.168	(0.092)	-0.225	(0.075)	0.274	(0.032)		
G026204	GI	8	0.721	(0.025)	0.478	(0.026)	0.000	(0.000)	-0.354 0.354	(0.053) (0.056)
G026401	GI	10	1.241	(0.157)	2.000	(0.117)	0.133	(0.010)		
G026402	GI	11	0.917	(0.114)	1.101	(0.069)	0.209	(0.022)		
G026501	GI	12	1.746	(0.136)	-0.773	(0.059)	0.303	(0.034)		
G026502	GI	13	1.027	(0.045)	0.077	(0.031)	0.000	(0.000)		
G026601	GI	15	0.916	(0.044)	1.332	(0.038)	0.000	(0.000)	0.191 -0.191	(0.040) (0.062)
G026701	GI	16	0.850	(0.067)	-0.734	(0.119)	0.255	(0.046)		
G026702	GI	17	1.152	(0.111)	0.636	(0.055)	0.224	(0.022)		
G026801	GI	18	1.091	(0.080)	-0.739	(0.088)	0.262	(0.040)		
G026802	GI	19	1.348	(0.118)	0.803	(0.040)	0.150	(0.016)		
G026803	GI	20	0.823	(0.134)	1.108	(0.099)	0.364	(0.027)		
G026901	GI	21	0.658	(0.029)	0.684	(0.033)	0.000	(0.000)	0.379 -0.379	(0.048) (0.058)
G027001	GI	22	0.600	(0.050)	-1.364	(0.181)	0.254	(0.056)		

Table D-28

IRT Parameters for the Geography Main Samples
Environment and Society Scale, Age 17/Grade 12

NAEP ID	Block	Item	A	S.E.	B	S.E.	C	S.E.	D	S.E.
G016401	GE	7	0.537	(0.023)	-0.430	(0.038)	0.000	(0.000)	0.144	(0.070)
									-0.144	(0.059)
G016701	GE	11	0.608	(0.028)	0.564	(0.034)	0.000	(0.000)	0.612	(0.051)
									-0.612	(0.060)
G017001	GE	14	1.581	(0.114)	0.071	(0.042)	0.213	(0.021)		
G017002	GE	15	0.699	(0.077)	0.315	(0.110)	0.252	(0.035)		
G017401	GF	1	0.598	(0.069)	-0.590	(0.222)	0.355	(0.058)		
G018201	GF	10	0.787	(0.118)	0.898	(0.104)	0.355	(0.030)		
G018501	GF	13	0.633	(0.024)	0.833	(0.028)	0.000	(0.000)	0.855	(0.056)
									0.556	(0.056)
									-1.410	(0.091)
G020701	GC	4	0.810	(0.035)	0.635	(0.027)	0.000	(0.000)	0.307	(0.041)
									-0.307	(0.049)
G021001	GC	8	0.604	(0.029)	1.667	(0.043)	0.000	(0.000)	1.163	(0.051)
									0.136	(0.073)
									-1.299	(0.181)
G021602	GC	15	0.599	(0.033)	1.378	(0.054)	0.000	(0.000)	0.617	(0.052)
									-0.617	(0.088)
G021701	GC	16	1.083	(0.106)	0.647	(0.057)	0.210	(0.022)		
G021702	GC	17	1.287	(0.133)	0.510	(0.059)	0.327	(0.023)		
G022301	GD	6	0.716	(0.030)	0.080	(0.028)	0.000	(0.000)	0.386	(0.049)
									-0.386	(0.047)
G022601	GD	10	0.938	(0.168)	1.541	(0.109)	0.283	(0.020)		
G023101	GD	17	0.535	(0.019)	1.742	(0.045)	0.000	(0.000)	2.269	(0.056)
									-0.982	(0.097)
									-1.287	(0.250)
G023601	GG	5	0.778	(0.031)	0.333	(0.025)	0.000	(0.000)	0.029	(0.045)
									-0.029	(0.047)
G023801	GG	7	0.696	(0.078)	-0.365	(0.174)	0.362	(0.049)		
G023901	GG	8	1.502	(0.154)	0.080	(0.070)	0.488	(0.026)		
G025001	GH	6	0.754	(0.041)	1.475	(0.052)	0.000	(0.000)	0.295	(0.046)
									-0.295	(0.080)
G025101	GH	7	0.778	(0.028)	0.849	(0.024)	0.000	(0.000)	0.861	(0.046)
									0.531	(0.046)
									-1.393	(0.083)

Table D-28 (continued)

IRT Parameters for the Geography Main Samples
Environment and Society Scale, Age 17/Grade 12

NAEP ID	Block	Item	A	S.E.	B	S.E.	C	S.E.	D	S.E.
G025301	GH	10	1.153	(0.075)	2.053	(0.059)	0.000	(0.000)	0.407	(0.042)
									0.023	(0.097)
									-0.430	(0.248)
G025701	GH	16	1.288	(0.170)	1.472	(0.082)	0.269	(0.015)		
G026101	GI	4	0.908	(0.034)	-0.907	(0.032)	0.000	(0.000)	0.076	(0.056)
									-0.076	(0.042)
G026301	GI	9	0.637	(0.034)	1.844	(0.061)	0.000	(0.000)	0.716	(0.049)
									-1.366	(0.154)
									0.649	(0.216)
G026503	GI	14	0.661	(0.027)	0.794	(0.027)	0.000	(0.000)	0.502	(0.052)
									0.198	(0.059)
									-0.701	(0.076)

Table D-29

IRT Parameters for the Geography Main Samples
Spatial Dynamics and Connections Scale, Age 17/Grade 12

NAEP ID	Block	Item	A	S.E.	B	S.E.	C	S.E.	D	S.E.
G016302	GE	6	0.904	(0.043)	1.215	(0.035)	0.000	(0.000)	0.420 -0.420	(0.037) (0.060)
G016502	GE	9	0.697	(0.032)	0.818	(0.034)	0.000	(0.000)	0.519 -0.519	(0.045) (0.059)
G016601	GE	10	1.107	(0.121)	0.562	(0.068)	0.321	(0.025)		
G016801	GE	12	0.610	(0.117)	1.151	(0.144)	0.353	(0.036)		
G017101	GE	16	0.640	(0.025)	0.494	(0.025)	0.000	(0.000)	0.667 -0.581 -0.085	(0.053) (0.066) (0.073)
G017301	GE	18	0.807	(0.094)	0.019	(0.135)	0.385	(0.040)		
G017701	GF	5	0.350	(0.013)	-0.365	(0.047)	0.000	(0.000)	-1.660 1.660	(0.119) (0.113)
G017801	GF	6	0.892	(0.070)	-1.119	(0.129)	0.289	(0.051)		
G017901	GF	7	0.630	(0.076)	0.007	(0.169)	0.324	(0.046)		
G018101	GF	9	0.776	(0.033)	0.488	(0.027)	0.000	(0.000)	0.340 -0.340	(0.043) (0.048)
G020501	GC	1	1.057	(0.091)	-0.657	(0.110)	0.345	(0.042)		
G020802	GC	6	1.385	(0.111)	-1.037	(0.086)	0.309	(0.042)		
G021301	GC	11	0.945	(0.108)	0.644	(0.076)	0.280	(0.026)		
G021501	GC	13	1.130	(0.130)	0.603	(0.070)	0.358	(0.024)		
G021801	GD	1	0.843	(0.075)	-0.562	(0.131)	0.315	(0.045)		
G022101	GD	4	0.639	(0.032)	1.048	(0.041)	0.000	(0.000)	0.294 -0.294	(0.050) (0.069)
G022501	GD	9	0.419	(0.017)	0.201	(0.038)	0.000	(0.000)	1.391 0.451 -1.842	(0.094) (0.077) (0.099)
G022801	GD	12	1.761	(0.158)	0.787	(0.037)	0.256	(0.016)		
G023002	GD	15	1.121	(0.111)	0.208	(0.078)	0.345	(0.028)		
G023003	GD	16	0.750	(0.092)	0.334	(0.123)	0.318	(0.038)		
G023501	GG	4	0.820	(0.030)	0.689	(0.023)	0.000	(0.000)	1.001 0.259 -1.260	(0.044) (0.043) (0.069)
G024101	GG	10	0.332	(0.042)	-1.118	(0.360)	0.325	(0.061)		

Table D-29 (continued)

IRT Parameters for the Geography Main Samples
Spatial Dynamics and Connections Scale, Age 17/Grade 12

NAEP ID	Block	Item	A	S.E.	B	S.E.	C	S.E.	D	S.E.
G024201	GG	11	2.019	(0.192)	0.493	(0.041)	0.362	(0.019)		
G024801	GH	2	1.021	(0.099)	0.487	(0.066)	0.240	(0.025)		
G025201	GH	8	1.333	(0.162)	1.030	(0.054)	0.304	(0.018)		
G025202	GH	9	0.814	(0.033)	0.715	(0.030)	0.000	(0.000)	0.847	(0.039)
									-0.847	(0.053)
G025501	GH	13	1.555	(0.128)	0.483	(0.041)	0.214	(0.019)		
G025502	GH	14	0.906	(0.093)	0.205	(0.094)	0.294	(0.033)		
G025601	GH	15	0.882	(0.052)	1.677	(0.059)	0.000	(0.000)	0.043	(0.049)
									-0.043	(0.087)
G025801	GH	17	0.666	(0.031)	1.089	(0.044)	0.000	(0.000)	1.170	(0.054)
									-1.170	(0.086)

Table D-29 (continued)

IRT Parameters for the Geography Main Samples
Spatial Dynamics and Connections Scale, Age 17/Grade 12

NAEP ID	Block	Item	A	S.E.	B	S.E.	C	S.E.	D	S.E.
G018401	GF	12	1.173	(0.153)	0.887	(0.065)	0.357	(0.021)		
G020901	GC	7	0.772	(0.069)	-0.661	(0.144)	0.298	(0.049)		
G022001	GD	3	0.863	(0.112)	0.476	(0.106)	0.388	(0.032)		
G022901	GD	13	1.288	(0.173)	1.022	(0.061)	0.367	(0.019)		
G023001	GD	14	1.089	(0.098)	0.533	(0.055)	0.192	(0.022)		

Table D-30

IRT Parameters for the Reading Long-Term Trend Samples, Age 9

NAEP ID	A	S.E.	B	S.E.	C	S.E.	D	S.E.
N001101	0.434	(0.125)	2.494	(0.463)	0.391	(0.033)		
N001501	2.105	(0.202)	-0.663	(0.062)	0.355	(0.029)		
N001502	2.185	(0.173)	0.275	(0.033)	0.211	(0.018)		
N001503	1.753	(0.158)	-0.171	(0.056)	0.337	(0.026)		
N001504	2.525	(0.228)	0.237	(0.035)	0.317	(0.020)		
N001507	0.749	(0.100)	2.678	(0.271)	0.000	(0.000)		
N001601	0.810	(0.085)	0.081	(0.100)	0.281	(0.031)		
N001602	1.593	(0.161)	0.536	(0.048)	0.316	(0.021)		
N001603	0.964	(0.140)	1.024	(0.080)	0.295	(0.024)		
N001604	1.365	(0.154)	0.865	(0.050)	0.242	(0.019)		
N001802	2.151	(0.225)	1.734	(0.082)	0.206	(0.011)		
N002001	1.511	(0.159)	1.006	(0.044)	0.177	(0.015)		
N002002	1.253	(0.137)	0.817	(0.053)	0.236	(0.020)		
N002003	1.381	(0.132)	0.535	(0.049)	0.253	(0.021)		
N002101	1.284	(0.233)	1.920	(0.160)	0.231	(0.014)		
N002102	1.540	(0.296)	2.194	(0.200)	0.156	(0.011)		
N002401	1.514	(0.135)	0.731	(0.038)	0.144	(0.016)		
N002702	1.477	(0.162)	0.921	(0.045)	0.203	(0.017)		
N002801	2.334	(0.186)	0.193	(0.032)	0.188	(0.018)		
N002802	2.083	(0.176)	0.118	(0.039)	0.231	(0.021)		
N002804	0.529	(0.058)	2.085	(0.189)	0.000	(0.000)		
N003001	0.899	(0.199)	1.998	(0.204)	0.173	(0.018)		
N003002	0.341	(0.062)	1.043	(0.261)	0.227	(0.048)		
N003101	1.126	(0.115)	0.131	(0.075)	0.290	(0.028)		
N003102	1.911	(0.188)	0.764	(0.038)	0.194	(0.016)		
N003104	0.754	(0.116)	2.768	(0.325)	0.000	(0.000)		
N003701	0.970	(0.105)	-0.034	(0.099)	0.327	(0.033)		
N003702	1.607	(0.157)	0.516	(0.046)	0.245	(0.021)		
N003704	0.781	(0.081)	0.912	(0.084)	0.000	(0.000)		
N003801	0.760	(0.218)	2.288	(0.341)	0.316	(0.021)		
N003802	0.505	(0.078)	0.580	(0.167)	0.265	(0.042)		
N003803	0.856	(0.284)	2.744	(0.511)	0.233	(0.016)		
N004101	0.917	(0.091)	-0.365	(0.109)	0.317	(0.035)		
N004201	1.216	(0.167)	1.089	(0.064)	0.255	(0.020)		
N004202	0.625	(0.113)	1.163	(0.137)	0.299	(0.032)		
N004701	1.678	(0.137)	0.431	(0.038)	0.190	(0.018)		
N004702	0.763	(0.093)	0.127	(0.121)	0.342	(0.035)		
N004703	1.188	(0.106)	0.297	(0.057)	0.232	(0.023)		
N004801	1.191	(0.111)	-0.300	(0.084)	0.352	(0.030)		
N004901	1.543	(0.162)	1.090	(0.050)	0.243	(0.016)		
N005101	0.540	(0.049)	-2.232	(0.257)	0.281	(0.062)		
N008601	1.363	(0.107)	-0.403	(0.063)	0.224	(0.026)		
N008602	1.093	(0.094)	0.000	(0.066)	0.217	(0.026)		

BEST COPY AVAILABLE

Table D-30 (continued)

IRT Parameters for the Reading Long-Term Trend Samples, Age 9

NAEP ID	A	S.E.	B	S.E.	C	S.E.	D	S.E.
N008603	1.071	(0.084)	-0.381	(0.073)	0.190	(0.027)		
N008701	0.473	(0.050)	-3.845	(0.375)	0.282	(0.065)		
N008801	1.142	(0.100)	-1.023	(0.107)	0.318	(0.037)		
N008901	1.894	(0.162)	-0.161	(0.048)	0.286	(0.024)		
N008902	1.193	(0.109)	-0.266	(0.079)	0.313	(0.029)		
N009001	1.605	(0.139)	0.557	(0.039)	0.187	(0.018)		
N009002	1.426	(0.159)	0.858	(0.049)	0.239	(0.019)		
N009003	1.639	(0.163)	1.233	(0.054)	0.239	(0.014)		
N009004	1.848	(0.173)	0.460	(0.042)	0.277	(0.020)		
N009101	0.896	(0.079)	-0.886	(0.121)	0.265	(0.038)		
N009201	1.337	(0.121)	-0.863	(0.092)	0.322	(0.034)		
N009401	1.263	(0.102)	-0.680	(0.077)	0.237	(0.031)		
N009601	0.748	(0.062)	-1.564	(0.161)	0.225	(0.049)		
N009701	1.089	(0.106)	0.220	(0.070)	0.270	(0.026)		
N009702	1.683	(0.147)	0.243	(0.046)	0.271	(0.022)		
N009703	1.324	(0.145)	0.721	(0.053)	0.263	(0.021)		
N009704	1.294	(0.135)	0.815	(0.049)	0.195	(0.019)		
N009705	1.633	(0.141)	0.124	(0.049)	0.276	(0.023)		
N009801	1.078	(0.102)	-1.897	(0.154)	0.328	(0.053)		
N009901	0.868	(0.089)	-0.038	(0.099)	0.283	(0.032)		
N010002	1.224	(0.112)	-0.023	(0.069)	0.281	(0.027)		
N010003	1.563	(0.134)	0.150	(0.049)	0.253	(0.023)		
N010102	1.156	(0.136)	0.682	(0.064)	0.286	(0.024)		
N010103	1.602	(0.132)	-0.047	(0.050)	0.233	(0.024)		
N010201	0.819	(0.075)	-1.572	(0.175)	0.308	(0.053)		
N010301	0.540	(0.056)	-1.418	(0.253)	0.301	(0.059)		
N010401	0.618	(0.064)	-0.933	(0.189)	0.272	(0.049)		
N010402	1.195	(0.170)	1.090	(0.067)	0.256	(0.021)		
N010403	1.328	(0.179)	1.303	(0.072)	0.233	(0.017)		
N010801	0.966	(0.097)	0.356	(0.071)	0.228	(0.026)		
N010902	1.810	(0.180)	0.584	(0.042)	0.275	(0.020)		
N010903	2.170	(0.181)	0.308	(0.035)	0.213	(0.019)		
N010904	2.295	(0.204)	0.687	(0.034)	0.277	(0.017)		
N011001	1.340	(0.087)	0.121	(0.042)	0.287	(0.018)		
N011002	1.798	(0.128)	0.623	(0.029)	0.278	(0.014)		
N011003	2.145	(0.131)	-0.089	(0.030)	0.297	(0.017)		
N011004	1.991	(0.116)	0.317	(0.026)	0.232	(0.014)		
N011101	1.770	(0.106)	0.374	(0.028)	0.222	(0.014)		
N011201	1.086	(0.090)	0.646	(0.047)	0.282	(0.018)		
N011301	1.694	(0.107)	0.252	(0.033)	0.284	(0.016)		
N011302	1.249	(0.108)	0.642	(0.046)	0.338	(0.017)		
N011401	2.118	(0.144)	1.216	(0.038)	0.377	(0.011)		
N011402	0.713	(0.085)	0.961	(0.081)	0.279	(0.023)		

Table D-30 (continued)

IRT Parameters for the Reading Long-Term Trend Samples, Age 9

NAEP ID	A	S.E.	B	S.E.	C	S.E.	D	S.E.
N011403	1.705	(0.132)	1.281	(0.044)	0.291	(0.011)		
N011404	1.858	(0.125)	1.228	(0.037)	0.241	(0.010)		
N013201	1.625	(0.132)	0.104	(0.045)	0.248	(0.022)		
N013301	1.330	(0.124)	-0.612	(0.087)	0.394	(0.031)		
N013401	1.264	(0.121)	0.603	(0.050)	0.218	(0.020)		
N013402	1.742	(0.154)	0.088	(0.049)	0.335	(0.023)		
N013403	1.897	(0.172)	0.546	(0.038)	0.254	(0.018)		
N014001	1.468	(0.139)	0.038	(0.061)	0.329	(0.026)		
N014101	0.752	(0.082)	-0.147	(0.122)	0.256	(0.037)		
N014201	1.008	(0.094)	-0.189	(0.087)	0.314	(0.030)		
N014301	2.105	(0.167)	0.230	(0.035)	0.214	(0.019)		
N014302	1.363	(0.135)	0.518	(0.052)	0.262	(0.022)		
N014303	2.389	(0.209)	0.015	(0.038)	0.292	(0.022)		
N014501	0.581	(0.034)	-1.014	(0.074)	0.000	(0.000)		
N014502	0.578	(0.034)	-1.206	(0.080)	0.000	(0.000)		
N014503	0.779	(0.042)	-1.677	(0.077)	0.000	(0.000)		

Table D-31

IRT Parameters for the Reading Long-Term Trend Samples, Age 13

NAEP ID	A	S.E.	B	S.E.	C	S.E.	D	S.E.
N001101	0.170	(0.034)	1.517	(0.540)	0.305	(0.043)		
N001201	0.482	(0.114)	1.801	(0.254)	0.369	(0.038)		
N001202	1.395	(0.161)	0.879	(0.053)	0.272	(0.019)		
N001301	0.712	(0.095)	0.163	(0.155)	0.409	(0.041)		
N001302	0.561	(0.071)	-2.389	(0.370)	0.526	(0.070)		
N001303	1.074	(0.120)	0.651	(0.069)	0.277	(0.025)		
N001401	0.853	(0.089)	-0.049	(0.111)	0.299	(0.039)		
N001501	1.823	(0.170)	-1.768	(0.088)	0.263	(0.049)		
N001502	1.522	(0.114)	-0.737	(0.059)	0.204	(0.031)		
N001503	1.154	(0.094)	-1.311	(0.105)	0.243	(0.046)		
N001504	1.131	(0.087)	-0.836	(0.084)	0.221	(0.037)		
N001507	0.499	(0.047)	2.227	(0.183)	0.000	(0.000)		
N001601	0.342	(0.043)	-1.220	(0.355)	0.281	(0.064)		
N001602	0.808	(0.070)	-1.454	(0.159)	0.273	(0.056)		
N001603	0.568	(0.075)	-0.071	(0.199)	0.316	(0.052)		
N001604	0.940	(0.086)	-0.329	(0.099)	0.274	(0.038)		
N001701	0.812	(0.072)	-0.781	(0.130)	0.257	(0.046)		
N001702	0.528	(0.156)	3.238	(0.571)	0.238	(0.022)		
N001703	0.694	(0.069)	-0.196	(0.126)	0.230	(0.041)		
N001802	1.101	(0.117)	0.649	(0.064)	0.254	(0.024)		
N001901	0.853	(0.085)	0.087	(0.095)	0.235	(0.035)		
N002001	1.028	(0.082)	-0.181	(0.071)	0.193	(0.030)		
N002002	1.084	(0.098)	-0.194	(0.083)	0.298	(0.033)		
N002003	1.033	(0.081)	-0.776	(0.092)	0.224	(0.038)		
N002101	0.691	(0.096)	1.116	(0.103)	0.199	(0.029)		
N002102	1.117	(0.109)	0.811	(0.053)	0.148	(0.019)		
N002201	0.932	(0.000)	-0.414	(0.065)	0.237	(0.030)		
N002202	1.037	(0.108)	-0.539	(0.121)	0.397	(0.043)		
N002203	0.431	(0.045)	-2.245	(0.304)	0.269	(0.063)		
N002401	0.858	(0.067)	-0.910	(0.107)	0.170	(0.041)		
N002501	0.647	(0.080)	0.263	(0.137)	0.262	(0.042)		
N002701	0.943	(0.106)	0.529	(0.081)	0.282	(0.029)		
N002801	1.344	(0.112)	-1.153	(0.089)	0.245	(0.041)		
N002802	1.140	(0.098)	-1.530	(0.121)	0.260	(0.051)		
N002902	0.460	(0.050)	-1.696	(0.296)	0.295	(0.067)		
N002903	1.500	(0.122)	-0.806	(0.070)	0.264	(0.035)		
N002904	0.845	(0.073)	-0.349	(0.099)	0.214	(0.037)		
N002905	0.588	(0.064)	0.258	(0.129)	0.199	(0.038)		
N002906	1.335	(0.104)	-0.855	(0.075)	0.230	(0.035)		
N003001	0.760	(0.113)	1.344	(0.102)	0.188	(0.025)		
N003002	0.330	(0.041)	-0.175	(0.253)	0.177	(0.050)		
N003003	1.983	(0.202)	2.137	(0.097)	0.077	(0.007)		
N003101	0.934	(0.080)	-1.203	(0.130)	0.264	(0.050)		

Table D-31 (continued)

IRT Parameters for the Reading Long-Term Trend Samples, Age 13

NAEP ID	A	S.E.	B	S.E.	C	S.E.	D	S.E.
N003102	1.360	(0.106)	-0.527	(0.065)	0.230	(0.031)		
N003104	0.560	(0.044)	1.455	(0.104)	0.000	(0.000)		
N003201	0.850	(0.077)	-0.920	(0.137)	0.283	(0.049)		
N003202	0.904	(0.079)	0.064	(0.078)	0.192	(0.030)		
N003203	1.077	(0.098)	0.111	(0.072)	0.250	(0.029)		
N003204	0.711	(0.083)	0.445	(0.108)	0.228	(0.035)		
N003301	0.792	(0.072)	-0.607	(0.126)	0.249	(0.044)		
N003401	1.038	(0.086)	-0.313	(0.080)	0.213	(0.034)		
N003501	0.709	(0.075)	-0.464	(0.155)	0.289	(0.049)		
N003601	1.124	(0.100)	-1.094	(0.114)	0.292	(0.048)		
N003602	0.977	(0.079)	-0.335	(0.081)	0.194	(0.033)		
N003701	0.959	(0.085)	-0.809	(0.113)	0.258	(0.044)		
N003702	1.218	(0.116)	-0.091	(0.076)	0.302	(0.032)		
N003704	0.674	(0.045)	0.176	(0.053)	0.000	(0.000)		
N003801	0.376	(0.076)	1.534	(0.256)	0.240	(0.048)		
N003802	0.255	(0.035)	-1.337	(0.400)	0.205	(0.059)		
N003803	0.396	(0.106)	2.974	(0.486)	0.240	(0.033)		
N003901	1.034	(0.109)	-2.752	(0.180)	0.259	(0.061)		
N004002	0.418	(0.047)	-2.654	(0.341)	0.270	(0.063)		
N004101	0.938	(0.078)	-1.719	(0.141)	0.252	(0.055)		
N004201	0.765	(0.080)	-0.062	(0.120)	0.274	(0.040)		
N004202	0.532	(0.065)	-0.215	(0.204)	0.283	(0.053)		
N004301	1.076	(0.109)	0.266	(0.075)	0.287	(0.030)		
N004303	1.041	(0.058)	0.300	(0.035)	0.000	(0.000)		
N004401	1.206	(0.123)	-2.547	(0.144)	0.256	(0.060)		
N004402	0.832	(0.069)	-0.305	(0.093)	0.194	(0.035)		
N004403	1.299	(0.115)	-1.712	(0.115)	0.273	(0.054)		
N004501	0.590	(0.077)	0.322	(0.155)	0.278	(0.043)		
N004502	0.557	(0.055)	-1.017	(0.201)	0.257	(0.055)		
N004601	0.851	(0.091)	0.358	(0.089)	0.246	(0.032)		
N004602	1.001	(0.081)	-0.252	(0.078)	0.199	(0.032)		
N004603	1.294	(0.104)	-0.567	(0.073)	0.267	(0.033)		
N004605	0.716	(0.043)	-1.108	(0.071)	0.000	(0.000)		
N004701	1.412	(0.104)	-0.861	(0.065)	0.175	(0.032)		
N004702	0.756	(0.062)	-1.300	(0.143)	0.226	(0.049)		
N004703	0.816	(0.064)	-1.275	(0.126)	0.210	(0.046)		
N004801	1.069	(0.084)	-1.398	(0.109)	0.222	(0.046)		
N004901	0.802	(0.074)	-0.224	(0.105)	0.223	(0.038)		
N005002	0.370	(0.091)	2.270	(0.352)	0.274	(0.042)		
N005003	0.767	(0.149)	2.150	(0.199)	0.158	(0.018)		
N005101	0.505	(0.056)	-3.116	(0.331)	0.266	(0.063)		
N005201	0.702	(0.183)	1.468	(0.220)	0.601	(0.027)		
N005202	0.429	(0.057)	0.268	(0.208)	0.227	(0.048)		

BEST COPY AVAILABLE

Table D-31 (continued)

IRT Parameters for the Reading Long-Term Trend Samples, Age 13

NAEP ID	A	S.E.	B	S.E.	C	S.E.	D	S.E.
N005203	1.132	(0.242)	2.090	(0.187)	0.323	(0.015)		
N005301	1.101	(0.114)	0.068	(0.082)	0.311	(0.032)		
N005302	1.853	(0.181)	0.580	(0.040)	0.211	(0.018)		
N005303	0.814	(0.126)	1.007	(0.106)	0.279	(0.029)		
N005304	2.006	(0.180)	0.154	(0.041)	0.230	(0.022)		
N005305	1.333	(0.145)	-0.392	(0.093)	0.406	(0.037)		
N005403	1.029	(0.093)	-0.742	(0.111)	0.304	(0.044)		
N005404	0.919	(0.081)	-1.825	(0.161)	0.287	(0.061)		
N005405	1.966	(0.170)	0.065	(0.044)	0.307	(0.023)		
N005406	0.902	(0.081)	-0.463	(0.107)	0.265	(0.041)		
N005407	1.194	(0.100)	-0.630	(0.084)	0.271	(0.038)		
N005503	0.883	(0.111)	0.489	(0.102)	0.358	(0.032)		
N005504	1.099	(0.127)	1.007	(0.062)	0.203	(0.020)		
N005505	0.678	(0.064)	-1.298	(0.193)	0.287	(0.060)		
N005601	1.390	(0.126)	-0.607	(0.082)	0.347	(0.037)		
N005602	1.253	(0.111)	0.461	(0.052)	0.198	(0.022)		
N005603	1.085	(0.100)	-0.531	(0.101)	0.329	(0.040)		

Table D-32

IRT Parameters for the Reading Long-Term Trend Samples, Age 17

NAEP ID	A	S.E.	B	S.E.	C	S.E.	D	S.E.
N001301	0.958	(0.121)	-0.215	(0.141)	0.523	(0.038)		
N001302	0.558	(0.078)	-2.767	(0.441)	0.585	(0.069)		
N001303	0.961	(0.091)	-0.226	(0.096)	0.281	(0.036)		
N001401	0.633	(0.071)	-0.998	(0.215)	0.328	(0.057)		
N001501	1.599	(0.169)	-2.157	(0.120)	0.294	(0.056)		
N001502	1.352	(0.112)	-1.230	(0.087)	0.198	(0.036)		
N001503	1.207	(0.117)	-1.591	(0.136)	0.295	(0.052)		
N001504	1.108	(0.094)	-1.381	(0.113)	0.232	(0.043)		
N001507	0.497	(0.047)	1.615	(0.137)	0.000	(0.000)		
N001701	0.663	(0.067)	-1.437	(0.210)	0.300	(0.058)		
N001703	0.785	(0.087)	-0.719	(0.166)	0.352	(0.050)		
N001901	1.022	(0.100)	-0.933	(0.127)	0.325	(0.046)		
N001904	0.774	(0.049)	-1.130	(0.074)	0.000	(0.000)		
N002001	1.200	(0.105)	-0.440	(0.081)	0.287	(0.035)		
N002002	1.133	(0.103)	-0.706	(0.098)	0.301	(0.040)		
N002003	1.259	(0.121)	-1.032	(0.106)	0.326	(0.043)		
N002101	0.674	(0.082)	0.271	(0.121)	0.209	(0.040)		
N002102	1.502	(0.129)	0.110	(0.053)	0.224	(0.027)		
N002201	1.462	(0.150)	-0.853	(0.096)	0.408	(0.040)		
N002202	2.041	(0.246)	-0.651	(0.078)	0.538	(0.035)		
N002203	0.386	(0.053)	-3.792	(0.524)	0.307	(0.066)		
N002501	0.561	(0.075)	-0.200	(0.206)	0.314	(0.053)		
N002701	0.947	(0.093)	-0.106	(0.096)	0.265	(0.037)		
N002702	0.840	(0.070)	-1.039	(0.120)	0.179	(0.042)		
N002801	1.174	(0.115)	-1.983	(0.151)	0.295	(0.057)		
N002802	1.368	(0.137)	-1.920	(0.130)	0.280	(0.053)		
N002804	0.246	(0.034)	1.928	(0.276)	0.000	(0.000)		
N002902	0.726	(0.083)	-1.224	(0.220)	0.373	(0.061)		
N002903	1.274	(0.123)	-1.346	(0.117)	0.292	(0.046)		
N002904	1.027	(0.100)	-0.812	(0.121)	0.336	(0.044)		
N002905	0.599	(0.066)	-0.101	(0.143)	0.218	(0.042)		
N002906	1.589	(0.159)	-1.198	(0.093)	0.292	(0.041)		
N003001	1.016	(0.094)	0.352	(0.064)	0.157	(0.027)		
N003002	0.380	(0.057)	-0.150	(0.287)	0.241	(0.059)		
N003003	2.038	(0.148)	1.137	(0.036)	0.088	(0.010)		
N003101	0.854	(0.082)	-2.084	(0.196)	0.307	(0.062)		
N003102	1.247	(0.117)	-1.143	(0.107)	0.286	(0.044)		
N003104	0.714	(0.053)	1.021	(0.068)	0.000	(0.000)		
N003201	1.007	(0.100)	-1.704	(0.168)	0.321	(0.058)		
N003202	1.021	(0.104)	-0.776	(0.129)	0.383	(0.046)		
N003203	1.004	(0.087)	-0.379	(0.089)	0.237	(0.037)		
N003204	1.059	(0.093)	-0.869	(0.105)	0.273	(0.042)		
N003301	0.981	(0.085)	-1.252	(0.122)	0.242	(0.045)		

Table D-32 (continued)

IRT Parameters for the Reading Long-Term Trend Samples, Age 17

NAEP ID	A	S.E.	B	S.E.	C	S.E.	D	S.E.
N003501	0.674	(0.068)	-0.879	(0.169)	0.271	(0.050)		
N003601	1.215	(0.118)	-1.580	(0.133)	0.286	(0.051)		
N003602	1.268	(0.111)	-0.682	(0.085)	0.286	(0.039)		
N003701	0.673	(0.074)	-1.244	(0.218)	0.332	(0.066)		
N003702	1.856	(0.162)	-0.390	(0.057)	0.323	(0.032)		
N003704	0.774	(0.072)	-0.711	(0.092)	0.000	(0.000)		
N003801	0.689	(0.113)	0.802	(0.127)	0.289	(0.038)		
N003802	0.199	(0.034)	-2.169	(0.599)	0.226	(0.062)		
N003803	0.601	(0.157)	1.898	(0.258)	0.308	(0.032)		
N004201	1.073	(0.111)	-0.402	(0.109)	0.358	(0.041)		
N004202	0.672	(0.085)	-0.183	(0.167)	0.314	(0.048)		
N004301	0.894	(0.095)	-0.356	(0.119)	0.272	(0.042)		
N004303	0.632	(0.054)	0.204	(0.065)	0.000	(0.000)		
N004501	0.751	(0.081)	-0.410	(0.144)	0.294	(0.046)		
N004502	0.592	(0.060)	-1.638	(0.234)	0.295	(0.060)		
N004601	0.774	(0.078)	-0.136	(0.111)	0.231	(0.038)		
N004602	1.339	(0.110)	-0.644	(0.072)	0.229	(0.034)		
N004603	1.383	(0.126)	-0.946	(0.088)	0.274	(0.038)		
N004605	0.588	(0.065)	-1.153	(0.152)	0.000	(0.000)		
N004901	0.742	(0.075)	-0.857	(0.161)	0.293	(0.050)		
N005001	1.809	(0.189)	0.697	(0.040)	0.228	(0.019)		
N005002	1.087	(0.135)	0.654	(0.073)	0.291	(0.027)		
N005003	0.946	(0.128)	1.142	(0.076)	0.156	(0.022)		
N005201	1.183	(0.255)	0.715	(0.127)	0.680	(0.024)		
N005202	0.547	(0.077)	0.146	(0.184)	0.285	(0.048)		
N005203	0.740	(0.152)	1.253	(0.138)	0.356	(0.032)		
N005503	0.780	(0.093)	0.001	(0.130)	0.327	(0.041)		
N005504	1.318	(0.128)	0.251	(0.063)	0.295	(0.027)		
N005505	0.790	(0.079)	-2.043	(0.227)	0.342	(0.067)		
N015101	1.296	(0.153)	0.309	(0.078)	0.419	(0.029)		
N015102	2.587	(0.202)	0.025	(0.032)	0.210	(0.020)		
N015103	2.701	(0.210)	0.138	(0.029)	0.174	(0.018)		
N015104	2.170	(0.192)	-0.034	(0.046)	0.324	(0.026)		
N015201	0.649	(0.064)	-2.297	(0.245)	0.303	(0.064)		
N015502	1.253	(0.114)	-0.207	(0.077)	0.319	(0.034)		
N015503	1.366	(0.135)	0.229	(0.064)	0.311	(0.028)		
N015504	1.329	(0.114)	-0.373	(0.072)	0.290	(0.034)		
N015505	0.633	(0.067)	-0.739	(0.177)	0.273	(0.050)		
N015901	1.191	(0.139)	0.144	(0.091)	0.415	(0.033)		
N015902	1.329	(0.146)	0.225	(0.075)	0.372	(0.030)		
N015903	1.947	(0.197)	0.575	(0.039)	0.228	(0.019)		
N016001	0.852	(0.086)	-0.769	(0.139)	0.299	(0.047)		
N016002	0.901	(0.117)	0.399	(0.103)	0.324	(0.035)		

Table D-32 (continued)

IRT Parameters for the Reading Long-Term Trend Samples, Age 17

NAEP ID	A	S.E.	B	S.E.	C	S.E.	D	S.E.
N016003	0.963	(0.093)	-0.165	(0.095)	0.270	(0.036)		
N016004	1.089	(0.102)	-0.446	(0.096)	0.301	(0.039)		
N016005	1.582	(0.134)	-0.408	(0.062)	0.276	(0.033)		
N016006	0.948	(0.089)	-0.097	(0.088)	0.226	(0.034)		
N017001	1.272	(0.121)	-0.012	(0.074)	0.336	(0.031)		
N017002	1.934	(0.178)	0.518	(0.037)	0.220	(0.019)		
N017003	2.463	(0.169)	1.181	(0.037)	0.203	(0.012)		

Table D-34

IRT Parameters for the Mathematics Long-Term Trend Samples, Age 9

NAEP ID	A	S.E.	B	S.E.	C	S.E.	D	S.E.
N250301	0.923	(0.109)	1.199	(0.068)	0.339	(0.018)		
N250601	0.967	(0.060)	-1.622	(0.107)	0.236	(0.051)		
N250602	0.599	(0.037)	-1.651	(0.143)	0.175	(0.048)		
N250603	0.986	(0.054)	-0.099	(0.052)	0.135	(0.023)		
N250701	0.645	(0.038)	-1.386	(0.119)	0.151	(0.042)		
N250702	1.163	(0.073)	0.531	(0.037)	0.161	(0.016)		
N250703	1.049	(0.053)	-0.442	(0.052)	0.123	(0.024)		
N250901	0.502	(0.034)	-1.798	(0.176)	0.178	(0.051)		
N250902	1.023	(0.066)	0.496	(0.043)	0.151	(0.018)		
N250903	1.115	(0.060)	-0.105	(0.045)	0.134	(0.021)		
N251401	0.677	(0.038)	-0.834	(0.095)	0.139	(0.036)		
N252001	1.194	(0.116)	1.706	(0.068)	0.213	(0.010)		
N252101	0.732	(0.105)	1.794	(0.108)	0.210	(0.018)		
N257201	0.853	(0.053)	-0.672	(0.091)	0.239	(0.037)		
N257801	0.645	(0.052)	-1.109	(0.186)	0.341	(0.057)		
N258501	0.544	(0.085)	1.835	(0.132)	0.215	(0.025)		
N261401	0.489	(0.040)	-0.460	(0.170)	0.233	(0.046)		
N262201	0.685	(0.052)	-0.684	(0.140)	0.310	(0.045)		
N262401	0.758	(0.067)	0.375	(0.089)	0.292	(0.029)		
N262501	0.502	(0.058)	0.171	(0.192)	0.332	(0.046)		
N263401	0.776	(0.054)	-1.261	(0.142)	0.306	(0.054)		
N263402	0.852	(0.061)	-0.546	(0.106)	0.313	(0.040)		
N265401	0.429	(0.137)	4.056	(0.838)	0.298	(0.019)		
N266101	0.578	(0.082)	1.296	(0.111)	0.275	(0.029)		
N267001	0.886	(0.055)	-1.404	(0.113)	0.283	(0.048)		
N267601	1.241	(0.068)	-0.615	(0.056)	0.231	(0.028)		
N267602	1.064	(0.056)	-0.114	(0.046)	0.150	(0.021)		
N268201	1.238	(0.088)	0.635	(0.039)	0.230	(0.016)		
N269001	0.830	(0.175)	2.901	(0.300)	0.091	(0.009)		
N269101	0.586	(0.086)	1.625	(0.110)	0.214	(0.025)		
N270001	0.596	(0.026)	-0.605	(0.042)	0.000	(0.000)		
N270901	0.618	(0.038)	-2.890	(0.140)	0.000	(0.000)		
N271101	0.748	(0.029)	-0.090	(0.030)	0.000	(0.000)		
N272101	0.839	(0.055)	-0.926	(0.112)	0.299	(0.044)		
N272102	0.890	(0.052)	-0.329	(0.068)	0.167	(0.029)		
N272301	0.738	(0.051)	-2.579	(0.162)	0.185	(0.053)		
N272801	0.715	(0.045)	-2.060	(0.140)	0.190	(0.051)		
N273501	0.602	(0.051)	-0.941	(0.198)	0.353	(0.056)		
N275401	1.031	(0.037)	-0.803	(0.029)	0.000	(0.000)		
N276001	0.946	(0.035)	-0.861	(0.032)	0.000	(0.000)		
N276002	0.889	(0.036)	0.964	(0.034)	0.000	(0.000)		
N276101	0.967	(0.037)	-1.051	(0.036)	0.000	(0.000)		
N276601	1.041	(0.059)	-1.203	(0.083)	0.214	(0.042)		
N276801	0.730	(0.047)	-3.119	(0.147)	0.000	(0.000)		

Table D-34 (continued)

IRT Parameters fcr the Mathematics Long-Term Trend Samples, Age 9

NAEP ID	A	S.E.	B	S.E.	C	S.E.	D	S.E.
N276802	0.595	(0.032)	-2.316	(0.106)	0.000	(0.000)		
N276803	0.506	(0.024)	-0.058	(0.041)	0.000	(0.000)		
N277401	0.824	(0.048)	-1.949	(0.109)	0.158	(0.043)		
N277501	0.795	(0.030)	-0.552	(0.032)	0.000	(0.000)		
N277601	0.844	(0.033)	-1.004	(0.038)	0.000	(0.000)		
N277602	0.780	(0.029)	-0.078	(0.029)	0.000	(0.000)		
N277603	0.792	(0.029)	-0.203	(0.029)	0.000	(0.000)		
N284001	0.801	(0.031)	-0.800	(0.037)	0.000	(0.000)		
N284002	0.729	(0.042)	1.967	(0.087)	0.000	(0.000)		
N286101	0.882	(0.033)	-0.861	(0.035)	0.000	(0.000)		
N286102	0.868	(0.031)	0.036	(0.026)	0.000	(0.000)		

Table D-35

IRT Parameters for the Mathematics Long-Term Trend Samples, Age 13

NAEP ID	A	S.E.	B	S.E.	C	S.E.	D	S.E.
N250201	0.617	(0.047)	-1.441	(0.183)	0.340	(0.055)		
N250701	0.474	(0.040)	-4.067	(0.310)	0.131	(0.047)		
N250702	0.801	(0.042)	-1.406	(0.087)	0.142	(0.036)		
N250703	0.526	(0.037)	-3.037	(0.203)	0.121	(0.043)		
N250901	0.406	(0.033)	-3.276	(0.278)	0.173	(0.051)		
N250902	0.942	(0.048)	-0.803	(0.063)	0.137	(0.030)		
N250903	0.874	(0.055)	-1.965	(0.115)	0.177	(0.044)		
N252001	1.094	(0.082)	0.633	(0.047)	0.257	(0.018)		
N252101	0.879	(0.082)	0.418	(0.082)	0.359	(0.028)		
N252901	1.175	(0.055)	-0.097	(0.035)	0.095	(0.017)		
N253701	0.327	(0.033)	-0.557	(0.240)	0.371	(0.040)		
N254001	0.854	(0.050)	-0.613	(0.077)	0.183	(0.034)		
N254601	0.977	(0.074)	-1.780	(0.139)	0.414	(0.053)		
N254602	0.919	(0.078)	1.041	(0.054)	0.211	(0.017)		
N255701	1.072	(0.075)	0.955	(0.040)	0.140	(0.014)		
N256101	0.857	(0.036)	-1.399	(0.048)	0.000	(0.000)		
N256501	1.387	(0.097)	0.401	(0.040)	0.298	(0.018)		
N256801	1.542	(0.105)	0.428	(0.035)	0.289	(0.016)		
N257601	1.289	(0.043)	-0.541	(0.021)	0.000	(0.000)		
N258801	1.241	(0.102)	0.659	(0.048)	0.354	(0.017)		
N258802	1.667	(0.099)	0.366	(0.028)	0.223	(0.014)		
N258803	1.505	(0.102)	0.943	(0.031)	0.182	(0.011)		
N260101	1.406	(0.081)	-0.021	(0.039)	0.229	(0.020)		
N261001	0.795	(0.064)	0.451	(0.074)	0.258	(0.026)		
N261201	0.599	(0.097)	2.083	(0.149)	0.250	(0.021)		
N261301	0.487	(0.050)	1.136	(0.108)	0.155	(0.029)		
N261501	0.690	(0.045)	-0.900	(0.119)	0.239	(0.043)		
N261801	0.567	(0.043)	-0.227	(0.125)	0.218	(0.038)		
N262201	0.432	(0.036)	-1.816	(0.251)	0.289	(0.058)		
N262401	1.048	(0.066)	-0.457	(0.071)	0.283	(0.032)		
N262501	0.515	(0.039)	-1.206	(0.180)	0.266	(0.050)		
N263101	0.686	(0.028)	-0.503	(0.034)	0.000	(0.000)		
N263401	0.743	(0.059)	-2.650	(0.201)	0.294	(0.059)		
N263402	0.585	(0.043)	-2.321	(0.201)	0.272	(0.056)		
N263501	0.933	(0.046)	-0.145	(0.046)	0.095	(0.021)		
N264701	1.192	(0.072)	0.370	(0.038)	0.197	(0.017)		
N265201	0.681	(0.055)	-2.549	(0.218)	0.316	(0.061)		
N265202	0.786	(0.060)	-0.417	(0.113)	0.341	(0.039)		
N265901	0.799	(0.067)	0.800	(0.064)	0.221	(0.022)		
N265902	0.661	(0.076)	1.145	(0.093)	0.272	(0.026)		
N266101	0.838	(0.054)	-0.796	(0.098)	0.272	(0.040)		
N266801	0.669	(0.047)	-1.187	(0.143)	0.299	(0.048)		
N267201	1.031	(0.079)	-0.753	(0.108)	0.452	(0.040)		
N269001	1.185	(0.066)	0.032	(0.042)	0.167	(0.021)		

Table D-35 (continued)

IRT Parameters for the Mathematics Long-Term Trend Samples, Age 13

NAEP ID	A	S.E.	B	S.E.	C	S.E.	D	S.E.
N269101	0.959	(0.061)	-0.280	(0.072)	0.230	(0.032)		
N269201	0.897	(0.042)	1.694	(0.058)	0.000	(0.000)		
N269901	0.801	(0.058)	-0.344	(0.098)	0.291	(0.036)		
N270301	0.369	(0.029)	-2.282	(0.260)	0.158	(0.054)		
N270302	1.329	(0.100)	1.728	(0.051)	0.067	(0.006)		
N273901	1.884	(0.105)	0.059	(0.027)	0.230	(0.016)		
N274801	1.399	(0.109)	0.285	(0.049)	0.419	(0.019)		
N275001	0.874	(0.034)	0.791	(0.031)	0.000	(0.000)		
N275301	0.298	(0.028)	-2.925	(0.354)	0.191	(0.055)		
N276801	0.434	(0.041)	-5.041	(0.418)	0.000	(0.000)		
N276802	0.463	(0.037)	-4.358	(0.306)	0.000	(0.000)		
N276803	0.342	(0.023)	-2.391	(0.160)	0.000	(0.000)		
N277401	0.528	(0.043)	-3.695	(0.280)	0.186	(0.054)		
N277601	0.652	(0.038)	-2.878	(0.136)	0.000	(0.000)		
N277602	0.644	(0.031)	-1.628	(0.069)	0.000	(0.000)		
N277603	0.597	(0.031)	-2.144	(0.098)	0.000	(0.000)		
N277901	0.666	(0.041)	-3.228	(0.157)	0.000	(0.000)		
N277902	0.683	(0.042)	-3.255	(0.157)	0.000	(0.000)		
N277903	0.731	(0.039)	-2.474	(0.103)	0.000	(0.000)		
N278901	1.375	(0.081)	0.160	(0.037)	0.236	(0.018)		
N278902	1.203	(0.099)	0.896	(0.044)	0.284	(0.015)		
N278903	1.808	(0.118)	0.775	(0.027)	0.221	(0.011)		
N278904	0.749	(0.085)	1.514	(0.080)	0.210	(0.019)		
N281401	0.618	(0.087)	2.225	(0.146)	0.156	(0.017)		
N281901	0.918	(0.073)	-2.538	(0.171)	0.242	(0.055)		
N282201	1.187	(0.082)	0.454	(0.044)	0.286	(0.018)		
N282202	1.020	(0.073)	-0.311	(0.079)	0.347	(0.032)		
N283101	2.339	(0.113)	0.981	(0.022)	0.134	(0.007)		
N285701	0.720	(0.060)	-0.073	(0.113)	0.288	(0.038)		
N286201	1.049	(0.065)	-0.686	(0.078)	0.312	(0.035)		
N286301	1.377	(0.081)	0.362	(0.033)	0.205	(0.016)		
N286501	1.037	(0.075)	0.949	(0.042)	0.147	(0.015)		
N286502	1.201	(0.079)	1.028	(0.035)	0.108	(0.011)		
N286601	1.034	(0.035)	-0.291	(0.023)	0.000	(0.000)		
N286602	1.037	(0.035)	-0.399	(0.024)	0.000	(0.000)		
N286603	1.175	(0.040)	0.540	(0.021)	0.000	(0.000)		

BEST COPY AVAILABLE

Table D-36

IRT Parameters for the Mathematics Long-Term Trend Samples, Age 17

NAEP ID	A	S.E.	B	S.E.	C	S.E.	D	S.E.
N251101	1.207	(0.044)	0.979	(0.028)	0.000	(0.000)		
N251701	0.936	(0.057)	-0.253	(0.067)	0.192	(0.030)		
N253901	1.217	(0.070)	-0.410	(0.054)	0.234	(0.027)		
N253902	0.908	(0.093)	0.632	(0.085)	0.430	(0.024)		
N253903	1.111	(0.087)	0.532	(0.056)	0.327	(0.020)		
N253904	1.660	(0.127)	0.551	(0.039)	0.380	(0.015)		
N254001	0.881	(0.057)	-1.033	(0.106)	0.257	(0.046)		
N254301	0.919	(0.070)	0.133	(0.077)	0.295	(0.029)		
N254601	1.122	(0.084)	-2.365	(0.125)	0.233	(0.055)		
N254602	1.233	(0.069)	-0.349	(0.050)	0.210	(0.026)		
N255501	0.800	(0.063)	0.216	(0.085)	0.272	(0.030)		
N255601	1.853	(0.118)	1.536	(0.045)	0.342	(0.009)		
N255701	1.301	(0.068)	-0.953	(0.053)	0.173	(0.031)		
N255801	0.801	(0.037)	1.595	(0.057)	0.000	(0.000)		
N256001	1.165	(0.039)	-0.229	(0.021)	0.000	(0.000)		
N256101	0.838	(0.044)	-2.149	(0.081)	0.000	(0.000)		
N256801	1.158	(0.068)	-0.552	(0.061)	0.234	(0.031)		
N257101	0.440	(0.089)	2.448	(0.226)	0.269	(0.030)		
N258801	1.221	(0.082)	-0.396	(0.066)	0.319	(0.031)		
N258802	1.525	(0.089)	-0.436	(0.044)	0.246	(0.026)		
N258803	1.268	(0.081)	0.091	(0.047)	0.255	(0.022)		
N258804	0.761	(0.054)	-2.259	(0.161)	0.248	(0.059)		
N259001	1.046	(0.037)	-0.379	(0.024)	0.000	(0.000)		
N259901	0.904	(0.061)	-0.277	(0.081)	0.267	(0.033)		
N260101	1.295	(0.072)	-1.283	(0.062)	0.196	(0.036)		
N260601	1.408	(0.064)	-1.612	(0.040)	0.000	(0.000)		
N260801	1.356	(0.044)	0.118	(0.019)	0.000	(0.000)		
N260901	1.905	(0.100)	-0.124	(0.027)	0.179	(0.016)		
N261001	0.807	(0.050)	-0.511	(0.087)	0.198	(0.036)		
N261201	0.560	(0.046)	-0.029	(0.136)	0.221	(0.041)		
N261301	0.506	(0.038)	0.173	(0.116)	0.149	(0.034)		
N261501	0.721	(0.045)	-1.979	(0.133)	0.191	(0.050)		
N261601	0.845	(0.118)	1.668	(0.092)	0.341	(0.017)		
N261801	0.626	(0.040)	-1.299	(0.139)	0.217	(0.049)		
N262301	0.579	(0.045)	-1.289	(0.198)	0.281	(0.062)		
N262401	0.866	(0.058)	-1.538	(0.127)	0.281	(0.054)		
N262501	0.550	(0.045)	-1.451	(0.218)	0.360	(0.059)		
N262502	1.248	(0.116)	1.352	(0.049)	0.262	(0.012)		
N262601	0.707	(0.052)	0.288	(0.082)	0.182	(0.029)		
N263001	0.617	(0.027)	0.883	(0.045)	0.000	(0.000)		
N263101	0.655	(0.028)	-0.915	(0.044)	0.000	(0.000)		
N263201	0.725	(0.050)	-1.763	(0.155)	0.309	(0.056)		
N263202	0.674	(0.054)	-0.859	(0.163)	0.351	(0.052)		
N264301	0.824	(0.033)	1.040	(0.039)	0.000	(0.000)		

Table D-36 (continued)

IRT Parameters for the Mathematics Long-Term Trend Samples, Age 17

NAEP ID	A S.E.	B S.E.	C S.E.	D S.E.
N264701	1.298 (0.073)	-0.323 (0.047)	0.214 (0.025)	
N266501	0.739 (0.054)	-0.534 (0.120)	0.278 (0.043)	
N268801	1.280 (0.082)	1.173 (0.035)	0.090 (0.009)	
N268901	1.568 (0.084)	0.375 (0.028)	0.165 (0.013)	
N269001	1.486 (0.091)	-0.224 (0.045)	0.279 (0.024)	
N270301	0.833 (0.053)	-2.101 (0.125)	0.167 (0.051)	
N270302	1.213 (0.056)	-0.109 (0.033)	0.091 (0.016)	
N271301	1.133 (0.071)	-0.123 (0.057)	0.263 (0.026)	
N278501	0.825 (0.033)	-0.958 (0.038)	0.000 (0.000)	
N278502	0.869 (0.032)	-0.578 (0.030)	0.000 (0.000)	
N278503	0.676 (0.029)	-1.063 (0.047)	0.000 (0.000)	
N278901	0.825 (0.047)	-0.842 (0.086)	0.182 (0.037)	
N278902	0.970 (0.063)	-0.389 (0.077)	0.248 (0.034)	
N278903	1.308 (0.077)	-0.123 (0.045)	0.238 (0.023)	
N278905	0.659 (0.077)	1.172 (0.095)	0.268 (0.027)	
N280401	0.542 (0.026)	-1.173 (0.060)	0.000 (0.000)	
N281401	0.579 (0.068)	1.596 (0.100)	0.172 (0.023)	
N286001	0.663 (0.039)	-1.392 (0.112)	0.165 (0.042)	
N286002	1.058 (0.058)	-1.541 (0.079)	0.141 (0.041)	
N286301	1.049 (0.058)	-0.891 (0.069)	0.210 (0.035)	
N286302	1.095 (0.064)	-0.833 (0.070)	0.236 (0.035)	
N286501	1.360 (0.081)	-0.911 (0.061)	0.255 (0.035)	
N286502	1.333 (0.066)	-0.518 (0.040)	0.143 (0.023)	
N287101	1.096 (0.065)	-0.529 (0.064)	0.246 (0.031)	
N287102	1.023 (0.052)	-0.870 (0.061)	0.153 (0.031)	
N287301	0.646 (0.027)	0.182 (0.033)	0.000 (0.000)	
N287302	0.800 (0.032)	0.947 (0.038)	0.000 (0.000)	

Table D-37

IRT Parameters for the Science Long-Term Trend Samples, Age 9

NAEP ID	A	S.E.	B	S.E.	C	S.E.	D	S.E.
N400001	0.630	(0.053)	-1.455	(0.190)	0.389	(0.045)		
N400101	0.607	(0.132)	1.958	(0.205)	0.519	(0.020)		
N400102	0.543	(0.090)	1.483	(0.152)	0.436	(0.025)		
N400301	0.836	(0.078)	0.080	(0.094)	0.478	(0.024)		
N400401	0.971	(0.080)	-1.471	(0.134)	0.498	(0.034)		
N400402	1.657	(0.099)	-0.892	(0.047)	0.290	(0.022)		
N400403	0.547	(0.048)	-2.252	(0.282)	0.466	(0.055)		
N400404	1.326	(0.085)	-0.718	(0.059)	0.367	(0.023)		
N400405	0.658	(0.050)	-1.217	(0.147)	0.359	(0.037)		
N400501	0.610	(0.070)	0.536	(0.122)	0.430	(0.027)		
N400601	0.916	(0.069)	0.053	(0.067)	0.341	(0.023)		
N400701	1.287	(0.085)	0.368	(0.039)	0.273	(0.016)		
N400901	0.262	(0.044)	1.827	(0.297)	0.331	(0.034)		
N401001	0.533	(0.049)	0.387	(0.104)	0.226	(0.028)		
N401101	0.508	(0.110)	2.047	(0.229)	0.467	(0.024)		
N401201	0.951	(0.138)	1.924	(0.104)	0.249	(0.013)		
N401301	0.570	(0.066)	0.837	(0.111)	0.337	(0.026)		
N401501	0.750	(0.127)	1.502	(0.126)	0.529	(0.019)		
N401601	0.562	(0.052)	-1.223	(0.212)	0.407	(0.046)		
N401702	0.360	(0.094)	2.180	(0.375)	0.589	(0.029)		
N401703	0.419	(0.086)	1.498	(0.239)	0.534	(0.030)		
N401801	1.026	(0.092)	-0.044	(0.082)	0.540	(0.022)		
N401802	1.183	(0.106)	-0.257	(0.081)	0.586	(0.022)		
N401803	0.830	(0.087)	0.192	(0.105)	0.564	(0.023)		
N401804	0.651	(0.102)	1.375	(0.129)	0.490	(0.022)		
N401901	0.464	(0.083)	1.788	(0.183)	0.380	(0.027)		
N402001	0.694	(0.053)	-1.290	(0.150)	0.385	(0.038)		
N402002	0.670	(0.052)	-1.571	(0.172)	0.400	(0.041)		
N402005	0.676	(0.067)	-0.265	(0.138)	0.493	(0.030)		
N402101	0.810	(0.067)	-0.049	(0.088)	0.396	(0.025)		
N402201	0.275	(0.034)	0.236	(0.268)	0.318	(0.039)		
N402401	0.429	(0.107)	2.997	(0.411)	0.339	(0.023)		
N402501	0.960	(0.110)	1.672	(0.076)	0.220	(0.013)		
N402602	0.421	(0.000)	-0.128	(0.213)	0.600	(0.028)		
N402701	0.753	(0.089)	1.571	(0.089)	0.253	(0.017)		
N402801	1.505	(0.104)	1.603	(0.047)	0.180	(0.008)		
N402901	0.312	(0.096)	5.343	(1.194)	0.202	(0.018)		
N403001	0.318	(0.045)	-7.652	(0.969)	0.320	(0.068)		
N403101	0.430	(0.040)	-5.368	(0.435)	0.309	(0.066)		
N403201	0.420	(0.030)	-2.987	(0.278)	0.266	(0.058)		
N403202	0.278	(0.027)	-1.624	(0.332)	0.268	(0.050)		
N403301	0.472	(0.038)	-1.515	(0.217)	0.280	(0.048)		
N403401	0.456	(0.062)	0.603	(0.181)	0.409	(0.034)		
N403501	0.436	(0.055)	0.045	(0.219)	0.412	(0.040)		

Table D-37 (continued)

IRT Parameters for the Science Long-Term Trend Samples, Age 9

NAEP ID	A	S.E.	B	S.E.	C	S.E.	D	S.E.
N403502	0.494	(0.044)	-2.598	(0.325)	0.462	(0.060)		
N403503	0.315	(0.046)	0.301	(0.298)	0.468	(0.039)		
N403601	0.700	(0.063)	0.656	(0.077)	0.266	(0.023)		
N403701	3.450	(0.000)	-0.346	(0.020)	0.373	(0.015)		
N403702	3.304	(0.000)	-0.444	(0.024)	0.474	(0.017)		
N403703	4.050	(0.315)	-0.305	(0.024)	0.374	(0.016)		
N403801	0.541	(0.107)	1.875	(0.193)	0.471	(0.023)		
N403803	0.402	(0.047)	-1.182	(0.338)	0.461	(0.052)		
N403804	0.321	(0.039)	-0.485	(0.309)	0.418	(0.043)		
N403901	0.549	(0.045)	-0.496	(0.133)	0.255	(0.034)		
N404001	0.206	(0.029)	1.500	(0.303)	0.257	(0.032)		
N404201	0.536	(0.061)	1.343	(0.102)	0.201	(0.023)		

Table D-38

IRT Parameters for the Science Long-Term Trend Samples, Age 13

NAEP ID	A	S.E.	B	S.E.	C	S.E.	D	S.E.
N400201	0.382	(0.031)	-1.885	(0.265)	0.285	(0.055)		
N401201	1.021	(0.070)	0.138	(0.057)	0.295	(0.022)		
N404501	0.909	(0.055)	-2.341	(0.127)	0.240	(0.057)		
N404601	0.537	(0.058)	0.079	(0.169)	0.325	(0.043)		
N404701	0.670	(0.042)	-1.770	(0.161)	0.274	(0.056)		
N404702	0.548	(0.057)	0.174	(0.148)	0.288	(0.040)		
N404801	0.895	(0.071)	-1.859	(0.188)	0.502	(0.060)		
N404802	1.699	(0.107)	-0.266	(0.042)	0.374	(0.020)		
N404803	1.429	(0.100)	0.454	(0.039)	0.354	(0.016)		
N404901	0.700	(0.045)	-0.729	(0.110)	0.228	(0.038)		
N405001	0.835	(0.078)	0.544	(0.076)	0.358	(0.024)		
N405101	0.882	(0.075)	0.920	(0.053)	0.221	(0.018)		
N405201	0.562	(0.084)	0.960	(0.135)	0.384	(0.033)		
N405301	1.122	(0.115)	1.167	(0.052)	0.319	(0.015)		
N405401	1.069	(0.090)	1.125	(0.045)	0.204	(0.014)		
N405501	1.025	(0.078)	0.343	(0.057)	0.331	(0.021)		
N405601	0.991	(0.142)	1.471	(0.078)	0.379	(0.016)		
N405701	1.147	(0.070)	0.428	(0.038)	0.189	(0.016)		
N405801	1.243	(0.108)	1.132	(0.042)	0.244	(0.013)		
N405901	1.249	(0.114)	1.644	(0.063)	0.220	(0.010)		
N406001	1.279	(0.232)	2.622	(0.248)	0.153	(0.007)		
N406101	1.772	(0.152)	2.022	(0.081)	0.191	(0.007)		
N406201	1.177	(0.135)	2.225	(0.123)	0.112	(0.007)		
N406301	0.693	(0.136)	1.034	(0.159)	0.665	(0.023)		
N406302	0.349	(0.038)	-1.094	(0.301)	0.491	(0.043)		
N406303	1.017	(0.093)	0.272	(0.076)	0.469	(0.023)		
N406304	0.503	(0.100)	1.391	(0.178)	0.468	(0.033)		
N406401	0.806	(0.093)	0.179	(0.126)	0.562	(0.028)		
N406402	1.191	(0.090)	0.115	(0.058)	0.415	(0.021)		
N406403	0.815	(0.085)	-1.063	(0.211)	0.608	(0.046)		
N406404	1.230	(0.102)	-0.242	(0.077)	0.531	(0.024)		
N406405	1.185	(0.098)	-0.227	(0.078)	0.510	(0.025)		
N406501	0.858	(0.071)	0.669	(0.058)	0.231	(0.021)		
N406601	0.444	(0.041)	-0.892	(0.238)	0.300	(0.055)		
N406701	0.801	(0.076)	0.454	(0.083)	0.354	(0.026)		
N406801	0.744	(0.059)	-1.825	(0.205)	0.464	(0.060)		
N406802	0.514	(0.089)	0.851	(0.196)	0.507	(0.036)		
N406803	0.649	(0.054)	-0.796	(0.165)	0.375	(0.046)		
N406804	0.697	(0.054)	-1.174	(0.170)	0.397	(0.049)		
N406805	1.435	(0.159)	1.135	(0.058)	0.552	(0.013)		
N406806	0.350	(0.043)	0.221	(0.235)	0.429	(0.037)		
N406901	0.664	(0.067)	0.233	(0.117)	0.339	(0.034)		
N407001	0.387	(0.056)	0.903	(0.202)	0.293	(0.043)		
N407101	0.972	(0.130)	1.842	(0.095)	0.155	(0.012)		

Table D-38 (continued)

IRT Parameters for the Science Long-Term Trend Samples, Age 13

NAEP ID	A	S.E.	B	S.E.	C	S.E.	D	S.E.
N407201	0.772	(0.083)	0.803	(0.080)	0.339	(0.024)		
N407301	0.376	(0.048)	0.905	(0.185)	0.226	(0.040)		
N407302	0.884	(0.177)	1.822	(0.136)	0.420	(0.017)		
N407601	0.801	(0.099)	1.373	(0.076)	0.264	(0.019)		
N407701	0.695	(0.076)	1.266	(0.075)	0.193	(0.021)		
N407801	1.097	(0.153)	2.012	(0.124)	0.289	(0.011)		
N407901	0.536	(0.070)	0.972	(0.120)	0.267	(0.033)		
N408001	1.342	(0.096)	0.917	(0.035)	0.198	(0.013)		
N408201	1.143	(0.133)	1.986	(0.102)	0.193	(0.009)		
N408301	1.497	(0.114)	0.796	(0.037)	0.335	(0.014)		
N408302	0.678	(0.052)	-1.464	(0.192)	0.426	(0.054)		
N408303	0.700	(0.064)	-1.287	(0.223)	0.514	(0.054)		
N408304	0.956	(0.072)	-1.216	(0.136)	0.464	(0.044)		
N408401	0.313	(0.031)	-0.490	(0.236)	0.312	(0.040)		
N408501	0.793	(0.058)	-0.682	(0.119)	0.326	(0.040)		
N408502	0.505	(0.058)	1.043	(0.107)	0.195	(0.030)		
N408601	0.526	(0.040)	-0.877	(0.174)	0.204	(0.051)		
N408701	0.339	(0.043)	0.710	(0.217)	0.385	(0.035)		
N408801	0.411	(0.077)	1.186	(0.219)	0.386	(0.044)		
N408901	1.032	(0.114)	0.698	(0.075)	0.545	(0.019)		
N408902	0.948	(0.078)	-1.702	(0.180)	0.528	(0.056)		
N408903	0.877	(0.085)	0.591	(0.075)	0.413	(0.022)		
N408904	0.830	(0.134)	1.361	(0.099)	0.499	(0.019)		
N409001	0.565	(0.043)	-0.312	(0.130)	0.206	(0.039)		
N409101	0.802	(0.056)	-1.112	(0.134)	0.327	(0.047)		
N409102	0.853	(0.076)	0.422	(0.076)	0.364	(0.024)		
N409103	1.051	(0.148)	1.748	(0.097)	0.371	(0.012)		
N409201	0.683	(0.094)	1.101	(0.100)	0.405	(0.025)		
N409301	0.740	(0.051)	-0.138	(0.084)	0.204	(0.030)		
N409501	0.813	(0.084)	1.626	(0.075)	0.127	(0.013)		
N409601	1.425	(0.116)	1.171	(0.040)	0.297	(0.012)		
N409701	1.279	(0.115)	1.769	(0.069)	0.210	(0.009)		

Table D-39

IRT Parameters for the Science Long-Term Trend Samples, Age 17

NAEP ID	A	S.E.	B	S.E.	C	S.E.	D	S.E.
N400201	0.476	(0.037)	-2.624	(0.252)	0.261	(0.059)		
N401201	0.836	(0.057)	-0.419	(0.092)	0.301	(0.034)		
N404601	0.415	(0.038)	-1.332	(0.259)	0.310	(0.055)		
N405001	0.456	(0.052)	-0.120	(0.223)	0.322	(0.050)		
N405101	1.007	(0.078)	0.238	(0.065)	0.335	(0.024)		
N405201	0.864	(0.108)	0.840	(0.084)	0.441	(0.023)		
N405401	0.977	(0.077)	0.687	(0.049)	0.212	(0.019)		
N405501	0.687	(0.050)	-0.523	(0.118)	0.277	(0.038)		
N406001	1.026	(0.143)	1.786	(0.098)	0.234	(0.012)		
N406101	1.354	(0.110)	1.453	(0.049)	0.195	(0.010)		
N406201	1.689	(0.097)	1.410	(0.035)	0.107	(0.007)		
N406301	1.332	(0.138)	0.185	(0.078)	0.634	(0.020)		
N406302	0.275	(0.032)	-1.450	(0.355)	0.450	(0.042)		
N406303	0.847	(0.073)	-0.226	(0.112)	0.443	(0.033)		
N406304	0.517	(0.056)	-0.489	(0.234)	0.439	(0.049)		
N406401	1.100	(0.093)	-0.241	(0.091)	0.523	(0.028)		
N406402	1.074	(0.077)	-0.602	(0.089)	0.439	(0.033)		
N406403	1.105	(0.093)	-1.365	(0.133)	0.550	(0.043)		
N406404	1.104	(0.077)	-1.087	(0.098)	0.437	(0.038)		
N406405	1.069	(0.076)	-1.269	(0.108)	0.431	(0.042)		
N406601	0.406	(0.034)	-1.669	(0.262)	0.238	(0.059)		
N406801	0.629	(0.056)	-2.495	(0.264)	0.465	(0.062)		
N406802	0.243	(0.042)	1.796	(0.370)	0.469	(0.028)		
N406803	0.690	(0.056)	-1.220	(0.178)	0.444	(0.048)		
N406804	0.585	(0.050)	-2.004	(0.247)	0.449	(0.058)		
N406805	0.565	(0.074)	0.285	(0.185)	0.476	(0.038)		
N406806	0.392	(0.049)	0.177	(0.235)	0.456	(0.039)		
N406901	0.603	(0.061)	-0.251	(0.172)	0.374	(0.045)		
N407001	0.327	(0.030)	-0.937	(0.245)	0.225	(0.047)		
N407101	1.084	(0.096)	1.088	(0.044)	0.188	(0.015)		
N407201	0.507	(0.048)	-0.107	(0.161)	0.235	(0.044)		
N407301	0.232	(0.029)	1.155	(0.252)	0.236	(0.032)		
N407302	0.979	(0.134)	1.193	(0.071)	0.419	(0.018)		
N407401	0.404	(0.045)	-0.559	(0.257)	0.477	(0.042)		
N407403	0.619	(0.064)	-0.218	(0.173)	0.446	(0.041)		
N407404	0.681	(0.057)	-1.833	(0.208)	0.461	(0.055)		
N407701	0.667	(0.057)	0.758	(0.070)	0.161	(0.024)		
N408101	0.943	(0.108)	1.421	(0.064)	0.193	(0.015)		
N408301	0.864	(0.065)	-0.435	(0.103)	0.402	(0.033)		
N408302	0.614	(0.052)	-2.011	(0.238)	0.453	(0.058)		
N408303	0.735	(0.061)	-1.856	(0.200)	0.472	(0.054)		
N408304	0.833	(0.067)	-1.875	(0.175)	0.467	(0.053)		
N408601	0.363	(0.029)	-2.571	(0.269)	0.177	(0.053)		
N408801	0.902	(0.073)	-0.028	(0.088)	0.367	(0.030)		

Table D-39 (continued)

IRT Parameters for the Science Long-Term Trend Samples, Age 17

NAEP ID	A	S.E.	B	S.E.	C	S.E.	D	S.E.
N408901	0.934	(0.070)	-1.008	(0.121)	0.466	(0.039)		
N408902	1.124	(0.099)	-1.983	(0.151)	0.507	(0.051)		
N408903	0.725	(0.062)	0.037	(0.106)	0.351	(0.032)		
N408904	0.574	(0.059)	-0.097	(0.167)	0.391	(0.040)		
N409301	0.791	(0.052)	-1.401	(0.128)	0.268	(0.048)		
N409501	0.715	(0.070)	1.120	(0.064)	0.150	(0.020)		
N409901	0.844	(0.054)	-0.995	(0.104)	0.279	(0.040)		
N410003	0.250	(0.034)	-4.217	(0.707)	0.488	(0.061)		
N410004	0.526	(0.069)	-0.568	(0.294)	0.628	(0.043)		
N410101	0.708	(0.072)	-0.703	(0.195)	0.556	(0.043)		
N410102	0.332	(0.039)	-0.977	(0.313)	0.506	(0.041)		
N410103	0.391	(0.041)	-1.436	(0.312)	0.496	(0.048)		
N410201	0.845	(0.112)	1.499	(0.080)	0.242	(0.017)		
N410401	0.678	(0.136)	1.311	(0.140)	0.561	(0.024)		
N410501	0.316	(0.028)	-0.909	(0.224)	0.168	(0.044)		
N410601	2.172	(0.099)	1.101	(0.023)	0.147	(0.008)		
N410602	0.475	(0.048)	-2.795	(0.374)	0.484	(0.068)		
N410603	1.865	(0.135)	0.877	(0.033)	0.415	(0.012)		
N410604	0.417	(0.041)	-2.584	(0.365)	0.445	(0.064)		
N410701	0.910	(0.073)	0.641	(0.055)	0.223	(0.021)		
N410801	0.561	(0.084)	1.259	(0.113)	0.275	(0.031)		
N410901	1.125	(0.078)	0.991	(0.035)	0.109	(0.012)		
N411001	1.262	(0.115)	1.271	(0.043)	0.183	(0.012)		
N411101	0.598	(0.060)	0.151	(0.134)	0.286	(0.039)		
N411201	1.128	(0.089)	0.475	(0.052)	0.323	(0.020)		
N411301	1.054	(0.256)	2.875	(0.358)	0.152	(0.008)		
N411401	1.886	(0.115)	0.456	(0.027)	0.242	(0.013)		
N411501	1.401	(0.108)	1.227	(0.038)	0.188	(0.010)		
N411502	0.779	(0.060)	-0.972	(0.142)	0.396	(0.045)		
N411601	1.384	(0.102)	0.898	(0.033)	0.192	(0.013)		
N411701	1.307	(0.099)	0.949	(0.034)	0.178	(0.013)		
N411801	2.012	(0.117)	0.457	(0.024)	0.206	(0.012)		
N411901	2.025	(0.107)	1.019	(0.025)	0.216	(0.009)		
N412001	1.204	(0.128)	1.603	(0.070)	0.266	(0.011)		

Table D-40

IRT Parameters for the Writing Long-Term Trend Samples, Grade 4

NAEP ID	A	S.E.	B	S.E.	C	S.E.	D	S.E.
N000602	0.396	(0.026)	0.540	(0.081)	0.000	(0.000)	-2.016	(0.222)
							2.016	(0.232)
N000902	0.453	(0.043)	3.176	(0.104)	0.000	(0.000)	2.565	(0.125)
							1.672	(0.171)
							-4.237	(1.547)
N001002	0.394	(0.032)	3.207	(0.114)	0.000	(0.000)	4.608	(0.170)
							1.036	(0.195)
							-5.644	(2.357)
N014702	0.498	(0.042)	-0.451	(0.076)	0.000	(0.000)	1.015	(0.133)
							-1.015	(0.111)
N014802	0.594	(0.047)	2.106	(0.068)	0.000	(0.000)	2.411	(0.098)
							0.817	(0.118)
							-3.227	(0.750)

Table D-41

IRT Parameters for the Writing Long-Term Trend Samples, Grade 8

NAEP ID	A	S.E.	B	S.E.	C	S.E.	D	S.E.
N000302	0.565	(0.049)	2.321	(0.092)	0.000	(0.000)	1.790	(0.098)
							0.448	(0.155)
							-2.239	(0.623)
N000402	0.422	(0.025)	2.415	(0.101)	0.000	(0.000)	3.982	(0.138)
							-0.369	(0.184)
							-3.613	(1.015)
N000502	0.585	(0.032)	1.470	(0.074)	0.000	(0.000)	3.195	(0.113)
							-0.304	(0.115)
							-2.891	(0.490)
N000602	0.287	(0.020)	-1.355	(0.131)	0.000	(0.000)	-3.421	(0.356)
							3.421	(0.327)
N000902	0.282	(0.024)	2.544	(0.110)	0.000	(0.000)	3.272	(0.199)
							1.685	(0.203)
							-4.957	(0.803)
N001002	0.374	(0.026)	1.228	(0.110)	0.000	(0.000)	4.358	(0.245)
							1.215	(0.140)
							-5.573	(0.689)

Table D-42

IRT Parameters for the Writing Long-Term Trend Samples, Grade 11

NAEP ID	A	S.E.	B	S.E.	C	S.E.	D	S.E.
N000302	0.929	(0.062)	1.304	(0.046)	0.000	(0.000)	1.691	(0.064)
							0.151	(0.080)
							-1.842	(0.244)
N000402	0.581	(0.032)	1.336	(0.074)	0.000	(0.000)	3.079	(0.116)
							-0.364	(0.116)
							-2.714	(0.421)
N001002	0.164	(0.012)	1.746	(0.238)	0.000	(0.000)	6.020	(0.534)
							3.502	(0.339)
							-9.522	(0.987)
N018002	0.313	(0.021)	1.799	(0.111)	0.000	(0.000)	3.861	(0.202)
							0.342	(0.183)
							-4.203	(0.619)
N019002	0.258	(0.015)	1.797	(0.139)	0.000	(0.000)	1.967	(0.316)
							4.928	(0.256)
							-6.896	(0.708)
N021002	0.596	(0.047)	1.569	(0.071)	0.000	(0.000)	1.637	(0.093)
							0.038	(0.127)
							-1.675	(0.322)

Appendix E

ESTIMATION ERROR VARIANCE BY GENDER AND RACE/ETHNICITY

Table E-1

Estimation Error Variance
for the Main Reading for Literary Experience Scale

		Proportion of Variance Due to . . .	
	Total Variance	**Student Sampling**	**Latency of θ**
GRADE 4			
Total	1.10	0.91	0.09
Male	1.59	0.90	0.10
Female	1.44	0.84	0.16
White	1.48	0.88	0.12
Black	3.13	0.82	0.18
Hispanic	6.64	0.91	0.09
Asian American	25.78	0.85	0.15
Pacific Islander	26.44	0.78	0.22
American Indian	13.54	0.76	0.24
Other	8.68	0.13	0.87
GRADE 8			
Total	0.96	0.84	0.16
Male	1.50	0.82	0.18
Female	1.10	0.78	0.22
White	1.42	0.84	0.16
Black	3.37	0.80	0.20
Hispanic	2.35	0.60	0.40
Asian American	9.49	0.57	0.43
Pacific Islander	99.68	0.94	0.06
American Indian	14.48	0.80	0.20
Other	10.67	0.07	0.93
GRADE 12			
Total	0.86	0.67	0.33
Male	1.51	0.63	0.37
Female	1.11	0.62	0.38
White	0.92	0.64	0.36
Black	3.89	0.66	0.34
Hispanic	5.32	0.64	0.36
Asian American	17.86	0.64	0.36
Pacific Islander	22.89	0.43	0.57
American Indian	93.45	0.88	0.12
Other	19.78	0.03	0.97

Table E-2

Estimation Error Variance
for the Main Reading to Gain Information Scale

		Proportion of Variance Due to . . .	
	Total Variance	**Student Sampling**	**Latency of θ**
GRADE 4			
Total	0.97	0.90	0.10
Male	1.57	0.88	0.12
Female	1.27	0.86	0.14
White	1.41	0.90	0.10
Black	3.05	0.78	0.22
Hispanic	6.15	0.90	0.10
Asian American	43.30	0.91	0.09
Pacific Islander	29.82	0.79	0.21
American Indian	14.93	0.82	0.18
Other	6.76	0.15	0.85
GRADE 8			
Total	0.75	0.88	0.12
Male	1.16	0.82	0.18
Female	0.90	0.85	0.15
White	0.99	0.87	0.13
Black	3.06	0.88	0.12
Hispanic	3.87	0.89	0.11
Asian American	10.80	0.78	0.22
Pacific Islander	24.46	0.85	0.15
American Indian	12.65	0.82	0.18
Other	8.77	0.08	0.92
GRADE 12			
Total	0.34	0.76	0.24
Male	0.58	0.76	0.24
Female	0.57	0.70	0.30
White	0.40	0.74	0.26
Black	1.49	0.73	0.27
Hispanic	1.59	0.73	0.27
Asian American	7.05	0.75	0.25
Pacific Islander	7.44	0.41	0.59
American Indian	19.79	0.82	0.18
Other	12.70	0.02	0.98

Table E-3

Estimation Error Variance for the Main Reading to Perform a Task Scale

		Proportion of Variance Due to . . .	
	Total Variance	**Student Sampling**	**Latency of θ**
GRADE 8			
Total	0.94	0.83	0.17
Male	1.21	0.80	0.20
Female	1.48	0.80	0.20
White	1.19	0.81	0.19
Black	3.87	0.77	0.23
Hispanic	2.55	0.65	0.35
Asian American	9.50	0.63	0.37
Pacific Islander	63.59	0.90	0.10
American Indian	17.34	0.82	0.18
Other	12.58	0.06	0.94
GRADE 12			
Total	0.70	0.71	0.29
Male	1.24	0.68	0.32
Female	1.02	0.72	0.28
White	0.93	0.68	0.32
Black	3.30	0.67	0.33
Hispanic	3.56	0.64	0.36
Asian American	11.36	0.74	0.26
Pacific Islander	18.26	0.35	0.65
American Indian	26.50	0.67	0.33
Other	16.95	0.03	0.97

Table E-4

Estimation Error Variance
for the Main Reading Composite Scale

		Proportion of Variance Due to . . .	
	Total Variance	Student Sampling	Latency of θ
GRADE 4			
Total	0.90	0.94	0.06
Male	1.41	0.93	0.07
Female	1.14	0.91	0.09
White	1.24	0.94	0.06
Black	2.62	0.85	0.15
Hispanic	5.98	0.94	0.06
Asian American	30.22	0.94	0.06
Pacific Islander	23.59	0.85	0.15
American Indian	12.68	0.81	0.19
Other	6.13	0.16	0.84
GRADE 8			
Total	0.70	0.93	0.07
Male	1.04	0.90	0.10
Female	0.85	0.91	0.09
White	0.94	0.93	0.07
Black	2.76	0.90	0.10
Hispanic	2.09	0.85	0.15
Asian American	7.51	0.78	0.22
Pacific Islander	53.08	0.96	0.04
American Indian	11.81	0.87	0.13
Other	7.26	0.09	0.91
GRADE 12			
Total	0.42	0.86	0.14
Male	0.71	0.86	0.14
Female	0.61	0.81	0.19
White	0.47	0.84	0.16
Black	1.97	0.82	0.18
Hispanic	2.27	0.80	0.20
Asian American	8.40	0.83	0.17
Pacific Islander	8.29	0.58	0.42
American Indian	30.20	0.89	0.11
Other	10.17	0.03	0.97

Table E-5

Estimation Error Variance
for the U.S. History Democracy Scale

		Proportion of Variance Due to . . .	
	Total Variance	**Student Sampling**	**Latency of θ**
GRADE 4			
Total	1.20	0.79	0.21
Male	2.34	0.79	0.21
Female	1.45	0.72	0.28
White	1.56	0.77	0.23
Black	3.84	0.68	0.32
Hispanic	8.64	0.81	0.19
Asian American	21.15	0.70	0.30
Pacific Islander	39.50	0.68	0.32
American Indian	39.79	0.90	0.10
Other	13.90	0.10	0.90
GRADE 8			
Total	0.53	0.84	0.16
Male	0.83	0.81	0.19
Female	0.62	0.76	0.24
White	0.80	0.83	0.17
Black	2.18	0.77	0.23
Hispanic	2.27	0.79	0.21
Asian American	14.90	0.88	0.12
Pacific Islander	73.72	0.95	0.05
American Indian	10.34	0.81	0.19
Other	5.50	0.09	0.91
GRADE 12			
Total	1.16	0.87	0.13
Male	1.47	0.81	0.19
Female	1.60	0.82	0.18
White	1.31	0.85	0.15
Black	4.05	0.83	0.17
Hispanic	4.44	0.74	0.26
Asian American	29.29	0.88	0.12
Pacific Islander	32.05	0.74	0.26
American Indian	27.06	0.81	0.19
Other	13.55	0.07	0.93

Table E-6

Estimation Error Variance
for the U.S. History Cultures Scale

		Proportion of Variance Due to . . .	
	Total Variance	Student Sampling	Latency of θ
GRADE 4			
Total	1.71	0.87	0.13
Male	3.30	0.85	0.15
Female	2.01	0.80	0.20
White	2.31	0.85	0.15
Black	4.27	0.71	0.29
Hispanic	10.54	0.85	0.15
Asian American	32.05	0.77	0.23
Pacific Islander	55.68	0.69	0.31
American Indian	51.07	0.85	0.15
Other	18.50	0.11	0.89
GRADE 8			
Total	0.45	0.85	0.15
Male	0.73	0.84	0.16
Female	0.60	0.75	0.25
White	0.76	0.85	0.15
Black	1.94	0.87	0.13
Hispanic	2.03	0.79	0.21
Asian American	12.92	0.84	0.16
Pacific Islander	44.12	0.92	0.08
American Indian	10.77	0.84	0.16
Other	4.43	0.09	0.91
GRADE 12			
Total	0.50	0.87	0.13
Male	0.68	0.80	0.20
Female	0.66	0.82	0.18
White	0.57	0.87	0.13
Black	1.45	0.76	0.24
Hispanic	2.25	0.78	0.22
Asian American	11.27	0.85	0.15
Pacific Islander	12.58	0.72	0.28
American Indian	12.54	0.83	0.17
Other	7.43	0.06	0.94

Table E-7

Estimation Error Variance
for the U.S. History Technology Scale

		Proportion of Variance Due to . . .	
	Total Variance	Student Sampling	Latency of θ
GRADE 4			
Total	1.56	0.88	0.12
Male	2.81	0.91	0.09
Female	1.97	0.78	0.22
White	2.07	0.88	0.12
Black	3.94	0.76	0.24
Hispanic	9.98	0.89	0.11
Asian American	32.87	0.84	0.16
Pacific Islander	61.30	0.73	0.27
American Indian	37.71	0.85	0.15
Other	15.01	0.14	0.86
GRADE 8			
Total	0.49	0.75	0.25
Male	0.75	0.71	0.29
Female	0.67	0.73	0.27
White	0.69	0.75	0.25
Black	1.89	0.74	0.26
Hispanic	2.78	0.79	0.21
Asian American	14.97	0.81	0.19
Pacific Islander	43.20	0.87	0.13
American Indian	9.76	0.81	0.19
Other	6.26	0.06	0.94
GRADE 12			
Total	0.46	0.86	0.14
Male	0.66	0.78	0.22
Female	0.58	0.85	0.15
White	0.53	0.82	0.18
Black	1.57	0.75	0.25
Hispanic	1.84	0.73	0.27
Asian American	13.09	0.85	0.15
Pacific Islander	13.97	0.72	0.28
American Indian	13.23	0.82	0.18
Other	5.86	0.07	0.93

Table E-8

Estimation Error Variance
for the U.S. History World Role Scale

		Proportion of Variance Due to . . .	
	Total Variance	Student Sampling	Latency of θ
GRADE 4			
Total	1.09	0.78	0.22
Male	1.93	0.73	0.27
Female	1.47	0.66	0.34
White	1.50	0.75	0.25
Black	3.27	0.58	0.42
Hispanic	5.89	0.71	0.29
Asian American	17.77	0.42	0.58
Pacific Islander	38.90	0.59	0.41
American Indian	23.14	0.78	0.22
Other	17.08	0.07	0.93
GRADE 8			
Total	0.73	0.79	0.21
Male	1.17	0.71	0.29
Female	0.91	0.74	0.26
White	1.10	0.80	0.20
Black	3.55	0.80	0.20
Hispanic	3.05	0.72	0.28
Asian American	19.22	0.82	0.18
Pacific Islander	53.16	0.91	0.09
American Indian	16.53	0.83	0.17
Other	7.36	0.08	0.92
GRADE 12			
Total	0.81	0.83	0.17
Male	1.00	0.80	0.20
Female	1.03	0.77	0.23
White	1.00	0.81	0.19
Black	3.33	0.83	0.17
Hispanic	3.46	0.70	0.30
Asian American	18.75	0.89	0.11
Pacific Islander	25.08	0.72	0.28
American Indian	19.36	0.76	0.24
Other	12.97	0.05	0.95

Table E-9

Estimation Error Variance
for the U.S. History Composite Scale

		Proportion of Variance Due to . . .	
	Total Variance	**Student Sampling**	**Latency of θ**
GRADE 4			
Total	1.09	0.95	0.05
Male	2.18	0.94	0.06
Female	1.27	0.90	0.10
White	1.40	0.94	0.06
Black	2.71	0.82	0.18
Hispanic	7.77	0.93	0.07
Asian American	19.83	0.90	0.10
Pacific Islander	38.47	0.84	0.16
American Indian	34.72	0.90	0.10
Other	10.83	0.15	0.85
GRADE 8			
Total	0.38	0.94	0.06
Male	0.62	0.91	0.09
Female	0.47	0.90	0.10
White	0.59	0.94	0.06
Black	1.77	0.91	0.09
Hispanic	1.83	0.90	0.10
Asian American	12.74	0.95	0.05
Pacific Islander	50.32	0.97	0.03
American Indian	9.87	0.87	0.13
Other	3.33	0.12	0.88
GRADE 12			
Total	0.60	0.95	0.05
Male	0.75	0.91	0.09
Female	0.75	0.93	0.07
White	0.68	0.94	0.06
Black	2.01	0.91	0.09
Hispanic	2.19	0.87	0.13
Asian American	15.51	0.95	0.05
Pacific Islander	15.13	0.88	0.12
American Indian	14.32	0.85	0.15
Other	7.27	0.08	0.92

BEST COPY AVAILABLE

Table E-10

Estimation Error Variance
for the Geography Space and Place Scale

		Proportion of Variance Due to . . .	
	Total Variance	Student Sampling	Latency of θ
GRADE 4			
Total	1.70	0.93	0.07
Male	2.40	0.93	0.07
Female	2.25	0.88	0.12
White	2.49	0.93	0.07
Black	7.22	0.90	0.10
Hispanic	6.72	0.87	0.13
Asian American	24.39	0.82	0.18
Pacific Islander	33.73	0.75	0.25
American Indian	12.91	0.74	0.26
Other	12.73	0.14	0.86
GRADE 8			
Total	0.71	0.88	0.12
Male	1.14	0.86	0.14
Female	0.88	0.85	0.15
White	0.93	0.86	0.14
Black	3.37	0.89	0.11
Hispanic	5.03	0.90	0.10
Asian American	8.13	0.74	0.26
Pacific Islander	74.53	0.94	0.06
American Indian	15.22	0.85	0.15
Other	6.04	0.11	0.89
GRADE 12			
Total	0.71	0.90	0.10
Male	0.97	0.87	0.13
Female	0.94	0.87	0.13
White	0.87	0.89	0.11
Black	2.54	0.85	0.15
Hispanic	2.98	0.84	0.16
Asian American	13.12	0.91	0.09
Pacific Islander	10.94	0.74	0.26
American Indian	9.46	0.68	0.32
Other	6.95	0.09	0.91

Table E-11

Estimation Error Variance
for the Geography Environment and Society Scale

		Proportion of Variance Due to . . .	
	Total Variance	Student Sampling	Latency of θ
GRADE 4			
Total	1.91	0.90	0.10
Male	2.41	0.85	0.15
Female	3.08	0.87	0.13
White	3.04	0.91	0.09
Black	6.98	0.85	0.15
Hispanic	9.30	0.83	0.17
Asian American	47.05	0.75	0.25
Pacific Islander	49.88	0.75	0.25
American Indian	16.95	0.70	0.30
Other	15.13	0.12	0.88
GRADE 8			
Total	0.50	0.85	0.15
Male	0.83	0.82	0.18
Female	0.65	0.79	0.21
White	0.62	0.80	0.20
Black	2.66	0.80	0.20
Hispanic	2.88	0.76	0.24
Asian American	7.61	0.64	0.36
Pacific Islander	79.62	0.95	0.05
American Indian	11.18	0.80	0.20
Other	4.78	0.09	0.91
GRADE 12			
Total	0.52	0.82	0.18
Male	0.75	0.77	0.23
Female	0.72	0.76	0.24
White	0.62	0.80	0.20
Black	1.69	0.83	0.17
Hispanic	2.48	0.76	0.24
Asian American	7.18	0.80	0.20
Pacific Islander	9.00	0.65	0.35
American Indian	7.06	0.72	0.28
Other	4.91	0.09	0.91

Table E-12

Estimation Error Variance
for the Geography Spatial Dynamics and Connections Scale

		Proportion of Variance Due to . . .	
	Total Variance	**Student Sampling**	**Latency of θ**
GRADE 4			
Total	1.63	0.85	0.15
Male	2.54	0.83	0.17
Female	2.10	0.82	0.18
White	2.50	0.87	0.13
Black	5.17	0.87	0.13
Hispanic	7.33	0.81	0.19
Asian American	26.41	0.79	0.21
Pacific Islander	33.47	0.66	0.34
American Indian	13.00	0.70	0.30
Other	13.88	0.11	0.89
GRADE 8			
Total	0.58	0.86	0.14
Male	0.96	0.81	0.19
Female	0.81	0.82	0.18
White	0.73	0.83	0.17
Black	3.11	0.84	0.16
Hispanic	4.18	0.87	0.13
Asian American	11.09	0.75	0.25
Pacific Islander	70.84	0.94	0.06
American Indian	11.51	0.84	0.16
Other	4.88	0.10	0.90
GRADE 12			
Total	0.63	0.86	0.14
Male	0.82	0.81	0.19
Female	0.87	0.84	0.16
White	0.71	0.84	0.16
Black	2.30	0.81	0.19
Hispanic	2.62	0.80	0.20
Asian American	13.79	0.89	0.11
Pacific Islander	14.63	0.75	0.25
American Indian	7.69	0.70	0.30
Other	6.28	0.08	0.92

Table E-13

Estimation Error Variance
for the Geography Composite Scale

		Proportion of Variance Due to . . .	
	Total Variance	Student Sampling	Latency of θ
GRADE 4			
Total	1.52	0.96	0.04
Male	2.11	0.95	0.05
Female	2.10	0.94	0.06
White	2.36	0.96	0.04
Black	5.79	0.94	0.06
Hispanic	6.28	0.91	0.09
Asian American	26.11	0.91	0.09
Pacific Islander	30.16	0.87	0.13
American Indian	11.79	0.76	0.24
Other	10.40	0.15	0.85
GRADE 8			
Total	0.51	0.96	0.04
Male	0.84	0.93	0.07
Female	0.66	0.92	0.08
White	0.64	0.94	0.06
Black	2.77	0.90	0.10
Hispanic	3.61	0.93	0.07
Asian American	6.74	0.88	0.12
Pacific Islander	71.68	0.98	0.02
American Indian	11.93	0.86	0.14
Other	4.20	0.12	0.88
GRADE 12			
Total	0.56	0.94	0.06
Male	0.75	0.92	0.08
Female	0.74	0.93	0.07
White	0.65	0.94	0.06
Black	1.94	0.91	0.09
Hispanic	2.26	0.90	0.10
Asian American	10.14	0.95	0.05
Pacific Islander	8.13	0.88	0.12
American Indian	7.04	0.75	0.25
Other	4.34	0.12	0.88

Table E-14

Estimation Error Variance
for the Reading Long-Term Trend Scale

		Proportion of Variance Due to . . .	
	Total Variance	**Student Sampling**	**Latency of θ**
AGE 9			
Total	1.55	0.88	0.12
Male	2.07	0.81	0.19
Female	2.18	0.87	0.13
White	1.75	0.86	0.14
Black	6.28	0.73	0.27
Hispanic	12.11	0.92	0.08
Asian American	35.34	0.88	0.12
American Indian	120.35	0.97	0.03
Other	265.92	0.63	0.37
AGE 13			
Total	0.85	0.83	0.17
Male	1.48	0.79	0.21
Female	1.30	0.82	0.18
White	1.31	0.85	0.15
Black	6.22	0.85	0.15
Hispanic	4.07	0.72	0.28
Asian American	23.31	0.83	0.17
American Indian	256.63	0.98	0.02
Other	210.05	0.77	0.23
AGE 17			
Total	2.01	0.87	0.13
Male	4.69	0.89	0.11
Female	2.15	0.82	0.18
White	2.29	0.86	0.14
Black	11.54	0.83	0.17
Hispanic	26.03	0.91	0.09
Asian American	53.99	0.87	0.13
American Indian	52.84	0.85	0.15
Other	211.41	0.92	0.08

Table E-15

Estimation Error Variance for the Mathematics Long-Term Trend Scale

	Total Variance	Proportion of Variance Due to . . . Student Sampling	Latency of θ
AGE 9			
Total	0.63	0.87	0.13
Male	0.95	0.83	0.17
Female	0.86	0.80	0.20
White	0.82	0.85	0.15
Black	3.00	0.79	0.21
Hispanic	4.44	0.90	0.10
Asian American	28.49	0.79	0.21
American Indian	34.04	0.93	0.07
Other	103.91	0.75	0.25
AGE 13			
Total	0.99	0.95	0.05
Male	1.49	0.92	0.08
Female	1.14	0.91	0.09
White	0.83	0.92	0.08
Black	12.86	0.96	0.04
Hispanic	2.95	0.86	0.14
Asian American	13.45	0.91	0.09
American Indian	29.63	0.89	0.11
Other	40.66	0.77	0.23
AGE 17			
Total	0.98	0.93	0.07
Male	1.85	0.91	0.09
Female	1.20	0.90	0.10
White	1.24	0.92	0.08
Black	3.29	0.87	0.13
Hispanic	13.57	0.95	0.05
Asian American	26.76	0.93	0.07
American Indian	49.51	0.92	0.08
Other	47.18	0.82	0.18

Table E-16

Estimation Error Variance for the Science Long-Term Trend Scale

		Proportion of Variance Due to . . .	
	Total Variance	**Student Sampling**	**Latency of θ**
AGE 9			
Total	1.01	0.84	0.16
Male	1.79	0.83	0.17
Female	1.19	0.70	0.30
White	1.19	0.81	0.19
Black	3.43	0.59	0.41
Hispanic	7.59	0.81	0.19
Asian American	33.91	0.66	0.34
American Indian	40.96	0.86	0.14
Other	61.01	0.36	0.64
AGE 13			
Total	1.08	0.90	0.10
Male	1.60	0.85	0.15
Female	1.58	0.86	0.14
White	0.90	0.83	0.17
Black	16.46	0.95	0.05
Hispanic	5.81	0.74	0.26
Asian American	25.66	0.89	0.11
American Indian	101.61	0.96	0.04
Other	38.66	0.57	0.43
AGE 17			
Total	2.39	0.94	0.06
Male	3.66	0.90	0.10
Female	3.05	0.93	0.07
White	2.38	0.92	0.08
Black	9.60	0.85	0.15
Hispanic	44.92	0.96	0.04
Asian American	82.81	0.95	0.05
American Indian	167.06	0.93	0.07
Other	137.76	0.85	0.15

Table E-17

Estimation Error Variance
for the Writing Long-Term Trend Scale

	Total Variance	Proportion of Variance Due to . . . Student Sampling	Latency of θ
GRADE 4			
Total	1.61	0.48	0.52
Male	2.31	0.48	0.52
Female	2.34	0.40	0.60
White	1.76	0.47	0.53
Black	7.03	0.43	0.57
Hispanic	6.68	0.56	0.44
Asian American	27.52	0.64	0.36
American Indian	23.60	0.61	0.39
Other	229.64	0.87	0.13
GRADE 8			
Total	1.08	0.59	0.41
Male	2.12	0.68	0.32
Female	1.47	0.48	0.52
White	1.13	0.52	0.48
Black	10.54	0.78	0.22
Hispanic	6.17	0.53	0.47
Asian American	18.00	0.72	0.28
American Indian	20.33	0.56	0.44
Other	78.42	0.87	0.13
GRADE 11			
Total	1.39	0.65	0.35
Male	2.06	0.63	0.37
Female	2.04	0.59	0.41
White	1.80	0.64	0.36
Black	6.76	0.66	0.34
Hispanic	11.85	0.73	0.27
Asian American	18.38	0.67	0.33
American Indian	26.67	0.60	0.40
Other	173.02	0.55	0.45

Appendix F

SETTING THE NAEP ACHIEVEMENT LEVELS FOR THE 1994 READING ASSESSMENT

Appendix F

SETTING THE NAEP ACHIEVEMENT LEVELS FOR THE 1994 READING ASSESSMENT

Mary Lyn Bourque
National Assessment Governing Board

INTRODUCTION

Since 1984, NAEP has reported the performance of students in the nation and for specific subpopulations on a 0-to-500 proficiency scale. The history and development of the scale and the anchoring procedure used to interpret specific points on that scale are described in Appendix G of *The NAEP 1992 Technical Report* (Johnson & Carlson, 1994).

The 1988 NAEP legislation[1] created an independent board, the National Assessment Governing Board (NAGB), responsible for setting policy for the NAEP program. The 1994 NAEP reauthorization[2] continued many of the Board's statutory responsibilities, including "developing appropriate student performance standards for each age and grade in each subject area to be tested under the National Assessment." Consistent with this directive, and striving to achieve one of the primary mandates of the statute "to improve the form and use of NAEP results," the Board has been developing student performance standards (called achievement levels by NAGB) on the National Assessment since 1990.

The 1990 standard-setting effort, initiated in December 1989 with the dissemination of a draft policy statement (NAGB, 1989) and culminating 22 months later in the publication of the NAGB report, *The Levels of Mathematics Achievement* (Bourque & Garrison, 1991), consisted of two phases: the main study and a replication-validation study. Although there were slight differences between the two phases, there were many common elements. Both phases used a modified (iterative/empirical) Angoff (1971) procedure for arriving at the levels; both focused on estimating performance levels based on a review of the 1990 NAEP mathematics item pool; and both phases employed policy definitions for basic, proficient, and advanced levels (NAGB, 1990) as the criteria for rating items. The 1990 process was evaluated by a number of different groups (for a discussion, see Hambleton & Bourque, 1991) who identified technical flaws in the 1990 process. These evaluations influenced the Board's decision to set the levels again in 1992, and to not use the 1990 levels as benchmarks for progress toward the national goals during the coming decade. It is interesting to note, however, that the 1990 and 1992 processes produced remarkably similar results.

---

[1] Public Law 100-297. (1988). National Assessment of Educational Progress improvement act (Article No. USC 1221). Washington, DC.

[2] Public Law 103-382. (1994). Improving America's schools act. Washington, DC.

In September 1991, the Board contracted with American College Testing (ACT) to convene the panels of judges that would recommend the levels on the 1992 NAEP assessments in reading, writing, and mathematics. While the 1992 level-setting activities were not unlike those undertaken by the Board in 1990, there were significant improvements made in the process for 1992. There was a concerted effort to bring greater technical expertise to the process: the contractor selected by the Board has a national reputation for setting standards in a large number of certification and licensure exams; an internal and external advisory team monitored all the technical decisions made by the contractor throughout the process; and state assessment directors periodically provided their expertise and technical assistance at key stages in the project.

Setting achievement levels is a method for setting standards on the NAEP assessment that identify what students should know and be able to do at various points along the proficiency scale. The initial policy definitions of the achievement levels were presented to panelists along with an illustrative framework for more in-depth development and operationalization of the levels. Panelists were asked to determine descriptions/definitions of the three levels from the specific framework developed for the NAEP assessment with respect to the content and skills to be assessed. The operationalized definitions were refined throughout the level-setting process, as well as validated with a supplementary group of judges subsequent to the level-setting meetings. Panelists were also asked to develop a list of illustrative tasks associated with each of the levels, after which sample items from the NAEP item pool were identified to exemplify the full range of performance of the intervals between levels. The emphasis in operationalizing the definitions and in identifying and selecting exemplar items and papers was to represent the full range of performance from the lower level to the next higher level. The details of the implementation procedures are outlined in the remainder of this appendix.

1992 PREPARATION FOR THE READING LEVEL-SETTING MEETING

It is important for the planning of any standard-setting effort to know how various process elements interact with each other. For example, panelists interact with pre-meeting materials, meeting materials (i.e., the assessment items, rating forms, rater feedback, and so forth), each other, and the project staff. All of these elements combine to promote or degrade what has been called intrajudge consistency and interjudge consensus (Friedman & Ho, 1990).

Previous research has conceptualized the effects of two major kinds of interaction: (1) people interacting with text (Smith & Smith, 1988), and (2) people interacting with each other (Curry, 1987; Fitzpatrick, 1989). In order to assess the effects of textual and social interaction and adjust the standard-setting procedures accordingly, a pilot study was conducted as the first phase of the 1992 initiative.

Reading was chosen as the single content area to be pilot tested since it combined all of the various features found in the other NAEP assessments, including multiple-choice, and both short and extended constructed-response items. The pilot study provided the opportunity to implement and evaluate all aspects of the operational plan—background materials, meeting materials, study design, meeting logistics, staff function, and participant function.

The overall pilot was quite successful. The level-setting process worked well, and the pilot allowed the contractor to make improvements in the design before implementation activities began. For example, schedule changes were made that allowed the panelists more time to operationalize the policy definitions before beginning the item-rating task. Also, the feedback mechanisms used to inform panelists about interjudge and intrajudge consistency data were improved for clarity and utility to the entire process.

1992 READING LEVEL-SETTING PANEL

Sixty-four panelists representing 32 jurisdictions (31 states and the Virgin Islands) were selected from the 366 nominees and invited to participate in the level-setting process. They represented reading/language arts teachers at grades 4, 8, and 12, nonteacher educators, and members of the noneducator (general public) community. The group was balanced by gender, race/ethnicity, NAEP regions of the country, community type (low SES, not low SES), district size, and school type (public/nonpublic). Two panelists were unable to attend due to a family emergency and a loss of job, resulting in 62 participants, 22 at grade 4, 20 at grade 8, and 20 at grade 12.

1992 PROCESS FOR DEVELOPING THE ACHIEVEMENT LEVELS

The four-and-one-half-day session began with a brief overview of NAEP and NAGB, a presentation on the policy definitions of the achievement levels, a review of the NAEP reading assessment framework, and a discussion of factors that influence item difficulty. The purpose of the presentation was to focus panelists' attention on the reading framework and to emphasize the fact that panelists' work was directly related to the NAEP assessment, not to the whole domain of reading.

All panelists completed and self-scored an appropriate grade-level form of the NAEP assessment. The purpose of this exercise was to familiarize panelists with the test content and scoring protocols—as well as time constraints—before beginning to develop the preliminary operationalized descriptions of the three levels.

Working in small groups of five or six, then eventually in grade-level groups, panelists expanded and operationalized the policy definitions of basic, proficient, and advanced in terms of specific reading skills, knowledge, and behaviors that were judged to be appropriate expectations for students in each grade, and to be in accordance with the current reading assessment framework.

The policy definitions[3] are as follows:

Basic	This level, below proficient, denotes partial mastery of the knowledge and skills that are fundamental for proficient work at each grade—4, 8, and 12.

---

[3] NAGB revised its policy definitions on achievement levels in late 1993. The *Proficient* level now reads: this level represents solid academic performance for each grade assessed. Students reaching this level have demonstrated competency over challenging subject matter, including subject-matter knowledge, application of such knowledge to real world situations, and analytical skills appropriate to the subject matter. *Basic* and *Advanced* remain virtually unchanged.

Proficient	This central level represents solid academic performance for each grade tested—4, 8, and 12. Students reaching this level have demonstrated competency over challenging subject matter and are well prepared for the next level of schooling.
Advanced	This higher level signifies superior performance beyond proficient grade-level mastery at grades 4, 8, and 12.

The small groups were allowed to brainstorm about what student performance *should be*, using the framework and their experience in completing the NAEP assessment as guides[4]. In addition, a practice task caused panelists to examine items in the half of the item pool that they would not be rating later. A comprehensive listing of grade level descriptors was developed, and panelists were asked to identify the five or six that best described what students *should be able to do* at each of the levels. Those descriptors appearing with the greatest frequency were compiled into a discussion list for the grade-level groups. Additions, deletions, and modifications were made as a result of discussions, and the groups reached general agreement that the final list of descriptors represented what students *should be able to do* at each achievement level.

Panelists next received training in the Angoff method, which was customized to reflect the unique item formats of the particular subject area assessment. Once a conceptual consensus was reached about the characteristics of **marginally** acceptable performance at each of the three levels, practice items from the released pool were rated by the panelists according to the process defined in the contractor's plan. For multiple-choice and short constructed-response items (both of which were scored right or wrong), panelists were asked to rate each item for the expected probability of a correct response for a group of *marginally* acceptable examinees at the basic, proficient, and advanced levels. For extended constructed-response items (which were scored on a four-point rating scale using a partial credit model), panelists were asked to review a set of student response papers and select three papers, one for each achievement level, that typified *marginally* acceptable examinee performance for that level.

Following training in the Angoff method, the judges began the rating and paper selection process, inspecting and rating each dichotomously scored item in the pool for the expected probabilities of answering the item correctly at each level. For polytomously scored items, panelists reviewed a representative set of 24 to 28 student response papers for each item and selected the paper that best represented marginally acceptable student performance at each level. Panelists completed three rounds of item ratings and paper selections. For Round 1, panelists first answered the items related to a reading passage, then reviewed their answers using scoring keys and protocols. This process helped ensure that panelists would be thoroughly familiar with each item, including the foils and scoring rubrics, before rating the item. Panelists provided item ratings/paper selections for all three achievement levels, one item at a time, for all the items related to a reading passage, then proceeded to the next reading passage and set of items, for which the process was repeated. Panelists rated items for half the items in their grade-level assessment; one block of exercises was common to both halves of the grade-level groups. During Round 1, panelists used their lists of descriptors and other training materials for guidance in the rating process.

---

[4] The panelists also reviewed about half the item pool (the half they would not be rating later) so that the descriptors could be further modified if that was deemed appropriate.

Following Round 1, item response theory (IRT) was used to convert the rating results[5] for each rater to a latent ability scale, represented by the Greek letter theta (θ). This θ scale was the same scale to which the NAEP items evaluated by each panelist were calibrated. In order to provide meaningful feedback about item ratings, a special *relative scale* was constructed, which was a linear transformation of the theta scale having a mean of 75 and standard deviation of 15. Before Round 2 of the rating process, panelists were given interjudge consistency information using this relative scale. This information allowed panelists to see where their individual mean item ratings were on the scale, relative to the mean for the group and to the means for other panelists. Reasons for extreme mean ratings, including the possibility that some panelists misinterpreted the item rating task, were discussed.

Before Round 2, panelists were also given item difficulty data. This information was presented as the overall percentage of students who answered each item correctly during the actual NAEP administration, for items scored "correct" or "incorrect" (i.e., multiple-choice and short constructed-response items), and as the mean score for student responses (on a scale of 1 to 4) for the extended response items. Panelists were told that this item difficulty information should be used as a reality check. For items on which item ratings differed substantially from the item difficulty value, panelists were asked to reexamine the item to determine if they had misinterpreted the item or misjudged its difficulty. Results of the data analysis, and panelists' own evaluations, indicated that the item difficulty information was perceived as very useful but had little impact on panelists' ratings.

For Round 2, panelists reviewed the same set of items they rated in Round 1 and, using the interjudge consistency information, the item difficulty information, and the information provided prior to Round 1, they either confirmed their initial item ratings and paper selections or adjusted their ratings to reflect the additional information. About one-half of Round 1 item ratings and paper selections were adjusted during Round 2.

Prior to Round 3, panelists' ratings were reanalyzed and additional information was presented to panelists concerning intrajudge variability. For each panelist, the intrajudge variability information consisted of those items that they had rated differently than items having similar difficulty, taking into consideration the panelist's aggregated item ratings. That is, the panelists' aggregated item ratings were converted to the theta (θ) scale. All items rated by the panelists were then analyzed in terms of the panelist's achievement level (θ) in comparison to actual student performance on the items. The observed item rating from each panelist was contrasted to an expected item rating. Those items with the largest differences between observed and expected ratings were identified. Panelists were given this information and asked to review each of these items and decide if their Round 2 ratings still accurately reflected their best judgments of the items. The intrajudge consistency data was to be used to flag items for reconsideration in the final round of rating.

For Round 3, panelists reviewed the same set of items they rated in Rounds 1 and 2 using both the new intrajudge variability information and the information made available during Rounds 1 and 2. In addition, panelists could discuss, within their small groups, ratings and paper selections for specific items about which they were unsure. About one-third of the item ratings were adjusted during Round 3.

---

[5] Because the IRT item parameters were not available for the polytomously scored (extended constructed-response) items, these items were not included in the following discussion of results.

1992 PROCESS FOR SELECTING EXEMPLAR ITEMS

On the final day of the achievement level-setting process, panelists reviewed items from the 1992 item pool scheduled for release to the public. The released item pool was the set from which the panelists could select items illustrative of the achievement levels for their grade. Exercises are organized in blocks, consisting of a reading passage, followed by several items, usually employing each of the three item formats, (i.e., multiple-choice, short constructed-response, and extended constructed-response). A total of 10 blocks from the 1992 exercise pool were scheduled for release: 2 blocks from the fourth-grade pool, totaling 19 items; 4 blocks from the eighth-grade pool, totaling 52 items; and 4 blocks from the twelfth-grade pool, totaling 46 items.

Panelists who had rated specific blocks of released items were asked to review those same items again to select particular ones as exemplary of each achievement level. The items were pre-assigned to each achievement level based on the final round of the judges' rating data, and using the following statistical criteria. For any given level (basic, proficient, or advanced),

1. items having an expected p-value[6] $\geq .501$ and $\leq .750$, *at that level*, were assigned to that level;

2. items meeting the criteria at *more than one level* were assigned to *one* level taking both the expected p-value and the appropriateness of the item for one of the levels into account; and

Because the content of items was given equal consideration in the selection process,

3. items with expected p-values $\leq .501$ were assigned to levels where a specific passage had few or no items at that level.

For example, the raters' expected p-value for one of the released items might have been .366 at the basic level, .701 at the proficient level, and .932 at the advanced level. This item would have been identified for review as a potential exemplar item for the proficient level. The expected p-value at the basic level was too low for consideration as a basic-level exemplar—that is, the item was judged to be too difficult, and the expected p-value at the advanced level was too high for consideration at the advanced level—that is, the item was judged to be too easy. Table F-1 shows the results of this process for each grade and level.

Panelists were asked to review the items as classified, and form an individual judgment regarding the suitability of each item to illustrate and further communicate the meaning of the levels. Each item's classification could be accepted, rejected, or reassigned, although the procedure was primarily designed to eliminate items that did not meet panelists' expectations for any reason. Items were reclassified if a strong consensus was found to hold for that change.

During the validation process, described in the next section, items were again reviewed. Those that had been selected by the original standard-setting panel were grouped into sets of *pre-selected* items. All remaining items in the released blocks that met the statistical criteria, *but were not recommended by the original panel*, were grouped into a set identified as *additional items for review*. Exercises that had been recommended for reclassification into another achievement level category were presented in their original classification for purposes of this review. As Table F-2 shows, 21 items were recommended as exemplars for the basic level, 17 for the proficient level, and 9 for the advanced.

---

[6] Expected p-values were based on the average predicted performance at the cut point for each achievement level.

Table F-1
Results of First Review for Achievement-Level Exemplars

Level/Status	Grade 4	Grade 8	Grade 12	All Grades
Total released	19	52	46	117
Basic				
Reviewed	4	12	18	34
Recommended	3	5	14	22
Proficient				
Reviewed	5	14	20	39
Recommended	4	12	9	25
Advanced				
Reviewed	5	6	7	18
Recommended	5	6	8	19

Table F-2
Results of Review of Additional Items for Achievement-Level Exemplars

Level/Status	Grade 4	Grade 8	Grade 12	All Grades
Total items recommended	13	13	21	47
Basic				
Reviewed	3	12	12	27
Recommended	6	7	8	21
Proficient				
Reviewed	4	13	11	28
Recommended	6	3	8	17
Advanced				
Reviewed	5	8	9	22
Recommended	1	3	5	9

BEST COPY AVAILABLE

1992 PROCESS FOR VALIDATING THE LEVELS

Nineteen reading educators participated in the item selection and content validation process. Ten of the panelists were reading teachers who had participated in the original achievement level-setting process and who had been identified as outstanding panelists by grade group facilitators during this meeting, who were extensively involved with professional organizations (e.g., the International Reading Association, the National Reading Conference, or the National Council for Teachers of English), and who had outstanding service credentials. The other nine panelists represented state-level reading curriculum supervisors or assessment directors, as well as university faculty teaching in disciplines related to this subject area. To the extent possible, the group was balanced by race/ethnicity and gender.

The two-and-one-half day meeting began by briefing panelists on the purpose of the meeting and by giving them an overview of the level-setting process and results. Panelists first reviewed the operationalized descriptions of the achievement levels for qualities such as (1) within- and across-grade consistency, (2) grade-level appropriateness, and (3) utility for increasing the public's understanding of the NAEP reading results. Next, panelists reviewed the operationalized descriptions of the achievement levels for consistency with the NAGB policy definitions of basic, proficient, and advanced with the NAEP *Reading Objectives*. Working in grade-level (4, 8, and 12) groups of 6 to 7 panelists each, then as a whole group, panelists reviewed the operationalized descriptions to provide within- and across-grade consistency, and to align the language and concepts of the descriptions more closely with the language of the NAEP *Reading Objectives*. (Both the original descriptions and the revised descriptions are included later in this appendix.) Finally, panelists suggested revisions they thought would improve the operational descriptions based on their earlier reviews.

On the final day, panelists worked in grade-level groups to review the possible exemplar items. The task was to select a set of items, for each achievement level for their grade, that would best communicate to the public the levels of reading ability and the types of skills needed to perform in reading at that level.

After selecting sets of items for their grades, the three grade-level groups met as a whole group to review item selection. During this process, cross-grade items that had been selected as exemplars for two grades (two such items were selected for grades 8 and 12) were assigned to one grade by whole-group consensus. In addition, items were evaluated by the whole group for overall quality. This process yielded 13 items as recommended exemplars for grade 4, 13 items as recommended exemplars for grade 8, and 21 items as recommended exemplars for grade 12.

EVALUATION OF THE 1992 LEVELS

The 1992 achievement levels in both mathematics and reading were evaluated under a Congressional mandate by the National Academy of Education (NAE). A series of research studies were mounted by the NAE (1993a; 1993b) to look at various aspects of the validity of the levels-setting process and the levels finally adopted by NAGB. Three of the studies focused specifically on the reading achievement levels, and were conducted for the Academy panel by staff at the Center for the Study of Reading at the University of Illinois at Urbana-Champaign. The first study examined the process for setting the levels in reading; the second study provided an analysis of the reading achievement levels descriptions; and the third focused on a comparison of the reading cut scores with those set by alternative means. Based on these studies the Academy's policy report concluded that the achievement levels were flawed and should be discontinued as a means of reporting NAEP data.

While NAGB did not agree with the conclusions reached in the NAE studies, and while the Board's technical advisors and contractor did not believe the weight of the evidence supported the conclusions reached by the Academy (American College Testing, 1993; Cizek, 1993; Kane, 1993), the Board agreed to support further investigation into the validity of the reading achievement levels through additional studies prior to the release of the 1994 NAEP reading data, since the Board planned on using the levels to report the 1994 NAEP data.

1994 PROCESS FOR VALIDATING THE LEVELS

The methodology developed by ACT to examine the reading achievement levels descriptions required the use of reading professionals (teachers and non-teacher educators) to review the descriptions in relation to the 1992 reading item pool. Fifty-eight panelists (about 20 at each grade level) were assigned to two different task groups, A and B. Group A employed the Item Difficulty Categorization (IDC) procedure, while Group B used a Judgmental Item Categorization (JIC) procedure. The goal of both task groups was to identify any lack of congruence between the item pool and the achievement level descriptions.

The IDC procedure examined the level of support for the descriptions as evidenced by performance on the NAEP items. Items were pre-selected for each achievement level using an response probability (rp) criterion of 0.50 at the lower borderline (*can do* items). Those items not meeting the same rp criterion at the upper borderline of the level were categorized as *can't do* items, while those items meeting the rp criterion anywhere in the range (from lower borderline to upper borderline) of the achievement level were labeled *challenging* items. Panelists were trained to examine the items in each of the three categories and determine whether or not the cognitive demand of the item matched the skills and knowledge identified in the descriptions. Mismatches were identified and later resolved or accounted for through a grade level procedure involving the JIC group.

The JIC procedure asked panelists to assign items to levels based on their judgment of where it belonged given the achievement levels descriptions. Items were assigned to the lowest level of performance required to respond correctly to the item. All items were assigned to levels independently by judges in the first round. Then, working in small groups and finally in the total group, assignments were confirmed and/or moderated through a consensus process.

The final grade-level procedure brought both groups A and B together to jointly evaluate the descriptions *vis a vis* performance on the item pool. The goal of the grade level procedure was to reach general agreement on the extent of (or lack of) agreement between the descriptions and the item pool, employing somewhat different approaches to the question.

On the basis of the validation process only one recommendation was made by the panelists to improve the descriptions and bring them more in line with the performance data they had examined during the process. The general conclusion was that reference to an ability to make inferences should be included in the description of Basic level achievement at each grade level. An adjustment has been made in the 1994 descriptions to reflect that recommendation.

1994 EXEMPLARS

The purpose of providing exemplar exercises is to provide readers with a sample of the kind of skills and knowledge that students reaching the achievement levels are likely to be able to respond to successfully. They are meant also to represent the kind of knowledge and skills embodied in the reading framework.

The selection of exemplar items for the 1994 reading assessment augment the 1992 exemplars by providing three additional passages (one for each grade level) and 13 additional exercises associated with the passages. The choice was made on the basis of criteria similar to those used in 1992, with one additional selection criterion, namely, item format. Since the percent of constructed response items increased by approximately 10 percent over the 1992 assessment, the choice of 1994 exemplars reflects this focus.

It should be noted that although some exemplars are associated with performance data from the 1992 and 1994 assessments (overall and conditional p-values), others have only 1992 performance estimates since they were released items in 1992 and not readministered in 1994. However, they are all reflective of the assessment framework.

MAPPING THE LEVELS ONTO THE NAEP SCALE

The process of mapping panelists' ratings to the NAEP scales used *item response theory* (IRT). IRT provided statistically sophisticated methods for determining the expected performance of examinees on particular test items in terms of an appropriate measurement scale. The same measurement scale simultaneously described the characteristics of the test items and the performance of the examinees. Once the item characteristics were set, it was possible to determine precisely how examinees were likely to perform on the test items at different points of the measurement scale.

The panelists' ratings of the NAEP test items were likewise linked, by definition, to the expected performance of examinees at the theoretical achievement level cut points. It was therefore feasible to use the IRT item characteristics to calculate the values on the measurement scale corresponding to each achievement level. This was done by averaging the item ratings over panelists for each achievement level and then simply using the item characteristics to find the corresponding achievement level cut points on the IRT measurement scale. This process was repeated for each of the NAEP reading scales within each grade (4, 8, and 12).

For the multiple-choice and short constructed-response items that were dichotomously scored, the judges each rated half of the items in the NAEP pool in terms of the expected probability that a student at a borderline achievement level would answer the item correctly, based on the judges' operationalization of the policy definitions and the factors that influence item difficulty. To assist the judges in generating consistently scaled ratings, the rating process was repeated twice, with feedback. Information on consistency among different judges and on the difficulty of each item[7] was fed back into the first repetition (Round 2), while information on consistency within each judge's set of ratings was fed back into the second repetition (Round 3). The third round of ratings permitted the judges to discuss their ratings among themselves to resolve problematic ratings. The mean final rating of the judges aggregated across multiple-choice and short constructed-response items yielded the threshold values for these items in the percent correct metric. These cut scores were then mapped onto the NAEP scale (which is defined and scored using item response theory, rather than percent correct).

---

[7] Item difficulty estimates were based on a preliminary, partial set of responses to the national assessment.

For extended constructed-response items, judges were asked to select student papers that exemplified performance at the cut point of each achievement level. Then for each achievement level, the mean of the scores assigned to the selected papers was mapped onto the NAEP scale in a manner similar to that used for the items scored dichotomously.

The final cut score for each achievement level was a weighted average of the cut score for the multiple-choice and short constructed-response items and the cut score for the extended constructed-response items, with the weights being proportional to the information supplied by the two classes of items. The judges' ratings, in both metrics, are shown for grade 4 in Table F-3.

Table F-3

Cut Points for Achievement Levels

Level	Mean Percent Correct, Multiple-Choice and Short Constructed-Response (Round 3)	Mean Paper Rating, Extended Constructed-Response (Round 3)	Scale Score*	Standard Error of Scale Score**
Grade 4				
Basic	38	2.72	208	(3.6)
Proficient	62	3.14	238	(1.4)
Advanced	80	3.48	268	(6.1)

***Note:** *Scale score is derived from a weighted average of the mean percents correct for multiple-choice and short constructed-response items and the mean paper ratings for extended constructed-response items after both were mapped onto the NAEP scale.*

****Note:** *The standard error of the scale is estimated from the difference in mean scale scores for the two equivalent subgroups of judges.*

In the final stage of the mapping process, the achievement level cut points on the IRT measurement scale were combined over content areas and rescaled to the NAEP score scale. Weighted averages of the achievement level cut points were computed. The weighting constants accounted for the measurement precision of the test items evaluated by the panelists, the proportion of items belonging to each NAEP content area, and the linear NAEP scale transformations. These weighted averages produced the final cut points for the basic, proficient, and advanced achievement levels within each grade.

BEST COPY AVAILABLE

Figure F-1

Final Descriptions of 1992 Reading Achievement Levels

PREAMBLE

Reading for meaning involves a dynamic, complex interaction between and among the reader, the text, and the context. Readers, for example, bring to the process their prior knowledge about the topic, their reasons for reading it, their individual reading skills and strategies, and their understanding of differences in text structures.

The texts used in the reading assessment are representative of common real world reading demands. Students at grade 4 are asked to respond to literary and informational texts which differ in structure, organization, and features. Literary texts include short stories, poems, and plays that engage the reader in a variety of ways, not the least of which is reading for fun. Informational texts include selections from textbooks, magazines, encyclopedias, and other written sources whose purpose is to increase the reader's knowledge.

In addition to literary and informational texts, students at grades 8 and 12 are asked to respond to practical texts (e.g., bus schedules or directions for building a model airplane) that describe how to perform a task.

The context of the reading situation includes the purposes for reading that the reader might use in building a meaning of the text. For example, in reading for literary experience, students may want to see how the author explores or uncovers experiences, or they may be looking for vicarious experience through the story's characters. On the other hand, the student's purpose in reading informational texts may be to learn about a topic (such as the Civil War or the oceans) or to accomplish a task (such as getting somewhere, completing a form, or building something).

The assessment asks students at all three grades to build, extend, and examine text meaning from four stances or orientations:

> **Initial Understanding**—Students are asked to provide the overall or general meaning of the selection. This includes summaries, main points, or themes.
>
> **Developing Interpretation**—Students are asked to extend the ideas in the text by making inferences and connections. This includes making connections between cause and effect, analyzing the motives of characters, and drawing conclusions.
>
> **Personal Response**—Students are asked to make explicit connections between the ideas in the text and their own background knowledge and experiences. This includes comparing story characters with themselves or people they know, for example, or indicating whether they found a passage useful or interesting.
>
> **Critical Stance**—Students are asked to consider how the author crafted a text. This includes identifying stylistic devices such as mood and tone.

These stances are not considered hierarchical or completely independent of each other. Rather, they provide a frame for generating questions and considering student performance at all levels. All students at all levels should be able to respond to reading selections from all of these orientations. What varies with students'

Figure F-1 (continued)

Final Descriptions of 1992 Reading Achievement Levels

developmental and achievement levels is the amount of prompting or support needed for response, the complexity of the texts to which they can respond, and the sophistication of their answers.

INTRODUCTION

The following achievement-level descriptions focus on the interaction of the reader, the text, and the context. They provide some specific examples of reading behaviors that should be familiar to most readers of this document. The specific examples are not inclusive; their purpose is to help clarify and differentiate what readers performing at each achievement level should be able to do. While a number of other reading achievement indicators exist at every level, space and efficiency preclude an exhaustive listing.

It should also be noted that the achievement levels are cumulative from basic to proficient to advanced. One level builds on the previous levels such that knowledge at the proficient level presumes mastery of the basic level, and knowledge at the advanced level presumes mastery at both the basic and proficient.

Grade 4—Basic

Fourth-grade students performing at the **basic level** *should demonstrate an understanding of the overall meaning of what they read.* When reading texts appropriate for fourth graders, *they should be able to make relatively obvious connections between the text and their own experiences*[8].

For example, when reading **literary text**, they should be able to tell what the story is generally about—providing details to support their understanding—and be able to connect aspects of the stories to their own experiences.

When reading **informational text**, basic-level fourth graders should be able to tell what the selection is generally about or identify the purpose for reading it; provide details to support their understanding; and connect ideas from the text to their background knowledge and experiences.

Grade 4—Proficient

Fourth grade students performing at the **proficient level** *should be able to demonstrate an overall understanding of the text, providing inferential as well as literal information.* When reading text appropriate

---

[8] Based on the recommendations of the 1994 reading revisit study, the phrase "*and extend the ideas in the text by making simple inferences*" has been added here to the description of *Basic*.

Figure F-1 (continued)

Final Descriptions of 1992 Reading Achievement Levels

to fourth grade, *they should be able to extend the ideas in the text by making inferences, drawing conclusions, and making connections to their own experiences. The connection between the text and what the student infers should be clear.*

For example, when reading **literary text**, proficient-level fourth graders should be able to summarize the story, draw conclusions about the characters or plot, and recognize relationships such as cause and effect.

When reading **informational text**, proficient-level students should be able to summarize the information and identify the author's intent or purpose. They should be able to draw reasonable conclusions from the text, recognize relationships such as cause and effect or similarities and differences, and identify the meaning of the selection's key concepts.

Grade 4—Advanced

Fourth-grade students performing at the **advanced level** *should be able to generalize about topics in the reading selection and demonstrate an awareness of how authors compose and use literary devices.* When reading text appropriate to fourth grade, *they should be able to judge texts critically and, in general, give thorough answers that indicate careful thought.*

For example, when reading **literary text**, advanced-level students should be able to make generalizations about the point of the story and extend its meaning by integrating personal experiences and other readings with the ideas suggested by the text. They should be able to identify literary devices such as figurative language.

When reading **informational text**, advanced-level fourth graders should be able to explain the author's intent by using supporting material from the text. They should be able to make critical judgments of the form and content of the text and explain their judgments clearly.

Grade 8—Basic

Eighth-grade students performing at the **basic level** *should demonstrate a literal understanding of what they read and be able to make some interpretations.* When reading text appropriate to eighth grade, *they should be able to identify specific aspects of the text that reflect the overall meaning,*[9] *recognize and relate interpretations and connections among ideas in the text to personal experience, and draw conclusions based on the text.*

For example, when reading **literary text**, basic-level eighth graders should be able to identify themes and make inferences and logical predictions about aspects such as plot and characters.

---

[9] Based on the recommendations of the 1994 reading revisit study, the phrase "*extend the ideas in the text by making simple inferences,*" has been added here to the description of *Basic*.

Figure F-1 (continued)
Final Descriptions of 1992 Reading Achievement Levels

When reading **informative text**, they should be able to identify the main idea and the author's purpose. They should make inferences and draw conclusions supported by information in the text. They should recognize the relationships among the facts, ideas, events, and concepts of the text (e.g., cause and effect and chronological order).

When reading **practical text**, they should be able to identify the main purpose and make predictions about the relatively obvious outcomes of procedures in the text.

Grade 8—Proficient

Eighth-grade students performing at the **proficient level** *should be able to show an overall understanding of the text, including inferential as well as literal information.* When reading text appropriate to eighth grade, *they should extend the ideas in the text by making clear inferences from it, by drawing conclusions, and by making connections to their own experiences—including other reading experiences.* Proficient eighth graders *should be able to identify some of the devices authors use in composing text.*

For example, when reading **literary text,** students at the proficient level should be able to give details and examples to support themes that they identify. They should be able to use implied as well as explicit information in articulating themes; to interpret the actions, behaviors, and motives of characters; and to identify the use of literary devices such as personification and foreshadowing.

When reading **informative text**, they should be able to summarize the text using explicit and implied information and support conclusions with inferences based on the text.

When reading **practical text**, proficient-level students should be able to describe its purpose and support their views with examples and details. They should be able to judge the importance of certain steps and procedures.

Grade 8—Advanced

Eighth-grade students performing at the **advanced level** *should be able to describe the more abstract themes and ideas of the overall text.* When reading text appropriate to eighth grade, they *should be able to analyze both meaning and form and support their analyses explicitly with examples from the text; they should be able to extend text information by relating it to their experiences and to world events.* At this level, student *responses should be thorough, thoughtful, and extensive.*

For example, when reading **literary text**, advanced-level eighth graders should be able to make complex, abstract summaries and theme statements. They should be able to describe the interactions of various literary elements (i.e., setting, plot, characters, and theme); to explain how the use of literary devices affects both the meaning of the text and their response to the author's style. They should be able critically to analyze and evaluate the composition of the text.

Figure F-1 (continued)

Final Descriptions of 1992 Reading Achievement Levels

When reading **informative text,** they should be able to analyze the author's purpose and point of view. They should be able to use cultural and historical background information to develop perspectives on the text and be able to apply text information to broad issues and world situations.

When reading **practical text**, advanced-level students should be able to synthesize information that will guide their performance, apply text information to new situations, and critique the usefulness of the form and content.

Grade 12—Basic

Twelfth-grade students performing at the **basic level** *should be able to demonstrate an overall understanding and make some interpretations of the text.* When reading text appropriate to twelfth grade, they *should be able to identify and relate aspects of the text to its overall meaning,*[10] *recognize interpretations, make connections among and relate ideas in the text to their personal experiences, and draw conclusions.* They *should be able to identify elements of an author's style.*

For example, when reading **literary text**, twelfth-grade students should be able to explain the theme, support their conclusions with information from the text, and make connections between aspects of the text and their own experiences.

When reading **informational text**, basic-level twelfth graders should be able to explain the main idea or purpose of a selection and use text information to support a conclusion or make a point. They should be able to make logical connections between the ideas in the text and their own background knowledge.

When reading **practical text**, they should be able to explain its purpose and the significance of specific details or steps.

Grade 12—Proficient

Twelfth-grade students performing at the **proficient level** *should be able to show an overall understanding of the text, which includes inferential as well as literal information.* When reading text appropriate to twelfth grade, they *should be able to extend the ideas of the text by making inferences, drawing conclusions, and making connections to their own personal experiences and other readings. Connections between inferences and the text should be clear, even when implicit.* These *students should be able to analyze the author's use of literary devices.*

When reading **literary text**, proficient-level twelfth graders should be able to integrate their personal experiences with ideas in the text to draw and support conclusions. They should be able to explain the author's use of literary devices such as irony or symbolism.

---

[10] Based on the recommendations of the 1994 reading revisit study, the phrase "*extend the ideas in the text by making simple inferences,*" has been added here to the description of *Basic*.

Figure F-1 (continued)

Final Descriptions of 1992 Reading Achievement Levels

When reading **informative text**, they should be able to apply text information appropriately to specific situations and integrate their background information with ideas in the text to draw and support conclusions.

When reading **practical texts**, they should be able to apply information or directions appropriately. They should be able to use personal experiences to evaluate the usefulness of text information.

Grade 12—Advanced

Twelfth-grade students performing at the **advanced level** *should be able to describe more abstract themes and ideas in the overall text.* When reading text appropriate to twelfth grade, they *should be able to analyze both the meaning and the form of the text and explicitly support their analyses with specific examples from the text.* They *should be able to extend the information from the text by relating it to their experiences and to the world.* Their *responses should be thorough, thoughtful, and extensive.*

For example, when reading **literary text**, advanced-level twelfth graders should be able to produce complex, abstract summaries and theme statements. They should be able to use cultural, historical, and personal information to develop and explain text perspectives and conclusions. They should be able to evaluate the text, applying knowledge gained from other texts.

When reading **informational text**, they should be able to analyze, synthesize, and evaluate points of view. They should be able to identify the relationship between the author's stance and elements of the text. They should be able to apply text information to new situations and to the process of forming new responses to problems or issues.

When reading **practical texts,** advanced-level twelfth graders should be able to make a critical evaluation of the usefulness of the text and apply directions from the text to new situations.

Figure F-2
Draft Descriptions of the Achievement Levels
Prepared by the Original Level-Setting Panel

4th-Grade Draft Descriptions

BASIC performance in reading should include:

- Determining what a text is about
- Identifying characterizations, settings, conflicts, or plots in a story
- Supporting one's understanding of a text with appropriate details
- Explaining why one likes or dislikes a text
- Connecting material in a text to personal experiences
- Making predictions about situations beyond the confines of a text
- Demonstrating an ability to maintain a focus over the entirety of a longer text

PROFICIENT performance in reading should include:

- Summarizing a text
- Recognizing an author's intent or purpose
- Making simple inferences based on information provided in a text
- Using information from a text to draw a basic conclusion
- Determining the meaning of key concepts in the text and connecting them to the main idea
- Recognizing the progression of ideas and the cause-and-effect relationships in a text
- Using the surrounding text to assign meaning to a word or phrase

ADVANCED performance in reading should include:

- Explaining an author's intent, using supporting material from the text
- Describing the similarities and differences in characters
- Demonstrating an awareness of the use of literary devices and figurative language
- Applying inferences drawn from a text to personal experiences
- Extending the meaning of a text by integrating experiences and information outside of the text
- Making and explaining a critical judgment of a text
- Demonstrating an ability to adapt reading purpose to genre and/or writing style

Figure F-2 (continued)

Draft Descriptions of the Achievement Levels
Prepared by the Original Level-Setting Panel

8th-Grade Draft Descriptions

BASIC performance in reading should include:

- Identifying the main idea or purpose of a text using information both stated and implied
- Expressing an author's purpose, viewpoint, and/or theme
- Using information from a text to draw and support conclusions
- Making inferences appropriate to the information provided in a text
- Recognizing the cause-and-effect relationships in a text
- Making logical connections from the material in a text to personal knowledge and experience

PROFICIENT performance in reading should include:

- Restating the main idea using supportive details and examples from a text
- Summarizing a text using information both stated and implied
- Making inferences from a text in order to draw valid conclusions
- Interpreting the actions, behaviors, and motives of characters
- Integrating personal knowledge and experience to enhance one's understanding of a text
- Identifying an author's use of literary devices

ADVANCED performance in reading should include:

- Describing how specific literary elements interact with each other
- Synthesizing the information in a text to obtain abstract meaning or to perform a task
- Finding new applications for information derived from a text
- Making personal and critical evaluations of a text
- Analyzing an author's purpose, viewpoint, and/or theme
- Explaining an author's use of literary devices

12th-Grade Draft Descriptions

BASIC performance in reading should include:

- Explaining the main idea of a text
- Describing the main purpose in reading a selection
- Recognizing the significance of details from a reading in order to support a conclusion or perform a task
- Applying the information gathered from reading to meet an objective or support a conclusion
- Explaining the basic elements of an author's literary devices

Figure F-2 (continued)
Draft Descriptions of the Achievement Levels
Prepared by the Original Level-Setting Panel

PROFICIENT performance in reading should include:

- Drawing conclusions from and making inferences about information from different texts and writing styles
- Integrating background information with newly acquired information to support conclusions
- Applying information from a text in an appropriate manner
- Bringing personal experience and accumulated knowledge into the process of critically evaluating a text
- Explaining an author's purpose in using complex literary devices

ADVANCED performance in reading should include:

- Providing innovative elaborations from textual information
- Analyzing and evaluating different points of view by means of comparison and contrast
- Identifying the relationships between an author's or narrator's stance and the various elements of the text
- Critically evaluating a text within a specific frame of reference
- Bringing the knowledge of other texts to the process of critical evaluation
- Using cultural or historical information provided in a text to develop perspectives on other situations
- Using cultural or historical information to develop perspectives on a text

Figure F-3

Revised Draft Descriptions of the Achievement Levels
Recommended by the Follow-Up Validation Panel

Revised 4th-Grade Draft Descriptions

BASIC performance in reading should include:

- Determining what a story/informational text is about (i.e., topic, main idea)
- Determining the main purpose for reading a selection
- Identifying character(s), setting(s), conflict(s), or plot(s) in a story
- Supporting one's understanding of a story/informational text with appropriate details
- Explaining why one likes or dislikes what they have read [a reading]
- Connecting material from a story/informational text to personal experiences
- Making predictions about situations beyond the confines of the printed material
- Maintaining a focus over the entirety of a story/informational text

PROFICIENT performance in reading should include:

- Summarizing a story/informational text
- Recognizing an author's intent or purpose
- Making simple inferences based on information provided in a story/informational text
- Drawing a valid conclusion from a story/informational text
- Determining the meaning of key concepts in the story/informational text and connecting them to the main idea
- Recognizing relationships in a story/informational text (i.e., time order, cause/effect, compare/contrast)

ADVANCED performance in reading should include:

- Explaining an author's intent, using supporting material from the story/informational text
- Describing the similarities and difference in characters, settings, and plots
- Demonstrating an awareness of the use of literary devices, such as figurative language
- Applying inferences drawn from a story/informational text to personal experiences
- Extending the meaning of a story/informational text by integrating experiences and information outside of the text
- Making and explaining a critical judgment of a story/informational text
- Demonstrating an ability to adapt reading purpose to a variety of printed material and/or writing style

Figure F-3 (continued)

Revised Draft Descriptions of the Achievement Levels
Recommended by the Follow-Up Validation Panel

Revised 8th-Grade Draft Descriptions

BASIC performance in reading should include:

- Identifying the main idea, theme, or purpose of a text
- Describing the main purpose for reading a selection
- Expressing an author's purpose and viewpoint
- Making inferences, predictions, and drawing conclusions that are supported by information in a text
- Recognizing the relationships among facts, ideas, events, and concepts within a text (i.e., cause and effect, chronological order, and characterization)
- Making logical connections between the text and personal knowledge
- Maintaining a focus over the entirety of a story/informational text

PROFICIENT performance in reading should include:

- Restating the main idea, theme, or purpose of a text using supporting details and examples
- Summarizing a text using both stated and implied information
- Interpreting the actions, behaviors, and motives of characters
- Using personal knowledge and experience to enhance one's understanding of a text
- Identifying an author's use of literary devices (i.e., personification, foreshadowing, and so forth)
- Using inferences from a text in order to draw valid conclusions

ADVANCED performance in reading should include:

- Describing how specific literary elements (i.e., setting, plot, characters, and theme) interact with each other
- Synthesizing the information in a text to obtain implied meaning or to perform a task
- Applying information derived from a text to new situations.
- Explaining an author's use of literary devices (i.e., irony, personification, and foreshadowing)
- Responding personally and critically to a text
- Analyzing an author's purpose and viewpoint
- Using cultural or historical information to develop perspectives on a text
- Using cultural or historical information provided in a text to develop perspectives on other situations

847

Figure F-3 (continued)

Revised Draft Descriptions of the Achievement Levels
Recommended by the Follow-Up Validation Panel

Revised 12th-Grade Draft Descriptions

BASIC performance in reading should include:

- Explaining the main idea, theme, or purpose of a text
- Describing the main purpose for reading a selection
- Recognizing the significance of details from a reading in order to support a conclusion or perform a task
- Applying the information gathered from reading to meet an objective or support a conclusion
- Identifying and explaining the basic elements of an author's literary devices
- Making logical connections between a text and personal knowledge and experience
- Maintaining a focus over the entirety of a story/informational text

PROFICIENT performance in reading should include::

- Drawing conclusions and making inferences from different texts and writing styles
- Integrating background information with newly acquired information to support conclusions
- Applying information from a text in an appropriate manner
- Applying personal experience and accumulated knowledge to the process of critically evaluating a text
- Explaining an author's purpose in using complex literary devices (i.e., irony, symbolism)

ADVANCED performance in reading should include:

- All basic and proficient reading behaviors listed previously
- Prompted by information from a text, innovating in new situations and creating new answers to old situations
- Analyzing, synthesizing, and evaluating different points of view by means of comparison and contrast
- Identifying the relationships between an author's or narrator's stance and the various elements of the text
- Critically evaluating a text within a frame of reference
- Applying the knowledge of other texts to the process of critical evaluation
- Using cultural or historical information to develop perspectives on a text
- Using cultural or historical information provided in a text to develop perspectives on other situations

Figure F-4

Meeting Participants, NAEP Reading Achievement-Level Setting
Original Meeting, St. Louis, Missouri, August 21 - 25, 1992

David Awbrey
Wichita Eagle
Wichita, KS

Dorothy Botham
Milwaukee Public Library
Milwaukee, WI

Anna Caballero
Attorney
Salinas, CA

Kathy Casseday
WFSP Radio Station
Kingwood, WV

Dee Ellis
Trimble Banner Newspaper
Milton, KY

Nona Smith
NAACP
New York, NY

Lillaine Speese
Oakdale Elementary School
Oroville, CA

Clifton Whetten
Retired Construction Sprvsr.
Elfrida, AZ

P. Richard Brackett
Brackett & Assoc. Motivational Marketing
Company
Brentwood, TN

Kathleen Harkey
Corporate Presentations
Nashville, TN

Patricia Oliverez
Salinas Public Library
Salinas, CA

Christine Sentz
North Milwaukee Branch Library
Milwaukee, WI

Carolyn Sullivan
Planters & Merchants Bank
Gillett, AR

Paula Abrams
City Hall
Bedford, KY

Rhonda Cantrell Dunn
Nashville Urban League
Nashville, TN

Harlon Gaskill (CPA)
Gaskill, Pharis & Pharis
Dalhart, TX

Jean McManis
Local/State Education Volunteer
State College, PA

Linda Borsum
Lakeview School District
Battlecreek, MI

Anne Kraut
Elementary Supervisor
Princeton, WV

Robert Williams
Macomb Intermediate SD
Clinton Township MI

Constance Boyd
Owen J. Roberts SD
King of Prussia, PA

Figure F-4 (continued)

Meeting Participants, NAEP Reading Achievement-Level Setting
Original Meeting, St. Louis, Missouri, August 21 - 25, 1992

Mary Gonzalez
Mesa Public Schools
Mesa, AZ

James Schindler
Jordan SD
Salt Lake City, UT

Kathryn Flannery
Indiana University
Bloomington, IN

Catherine Hatala
School District of Philadelphia
Philadelphia, PA

Raymond Morgan
Old Dominion University
Virginia Beach, VA

Berton Wiser
Columbus Public School
Columbus, OH

Freda Andrews
Durham Public Schools
Durham, NC

Tim Barnes
Ashdown Public Schools
Ashdown, AR

Larry Barretto
Maplewood Elementary School
Coral Springs, FL

Gloria Darling
Conway Public Schools
Conway, AR

Nina Frederick
Marion County School System
Hackleburg, AL

Karen Fugita
Oak Grove SD
San Jose, CA

Anne Gregory
Durham Public Schools
Durham, NC

Joseph Howard
Josiah Quincy School
West Roxbury, MA

Roberta Johnson
Cleveland Public Schools
Cleveland, OH

Marcia Jolicoeur
Lisbon Falls School
Lewiston, ME

Elizabeth Litchfield
Westwood School District
Emerson, NJ

Jean Young
Houston ISD
Houston, TX

Wilma Centers
Wolfe County Middle School
Campton, KY

Eunice Coakley
Greenville School
Greenville, SC

Eugenia Constantinou
Prince Georges County Schools
Silver Spring, MD

Walt Cottingham
Henderson City Schools
Zirconia, NC

Figure F-4 (continued)

Meeting Participants, NAEP Reading Achievement-Level Setting
Original Meeting, St. Louis, Missouri, August 21 - 25, 1992

Deborah Davidson
Westhampton Beach UFSD
Patchogue, NY

Julia Dominique
Department of Education USVI
Sunnyisle, VI

Patricia Gerdes
Waelder ISD
Schulenburg, TX

Leslie Leech
Elkton School
Elkton, SD

Belva Leffel
Whittier Christian Jr. High
Norwalk, CA

Harriett McAllaster
Volusia County Schools
DeLand, FL

Mary Orear
Camden-Rockport HS & MS
Rockport, ME

Judith Zinsser
Houston ISD
Houston, TX

Mary Ann Ledbetter
East Baton Rouge Parish School Board
Baton Rouge, LA

Cora Cummins
Conway Public Schools
Conway, AR

Stanley Fraundorf
Cuba City Public Schools
Cuba City, WI

Georgia Howard
Volusia County Schools
Holly Hill, FL

Roger Larsen
Campbell County SD
Gillette, WY

Judith Lusk
Norfield School District
Rockbury, VT

Donnie McQuinn
Wolfe County Board of Education
Pine Ridge, KY

Meredith Powers
Swansea School
Providence, RI

Beth Schieber
Kingfisher Schools
Okarche, OK

Carolyn Sue Wilson
Greenville, SC

Sue Zak
Cleveland Board of Education
Garfield Heights, OH

Figure F-5

Meeting Participants, NAEP Reading Achievement-Level Setting
Follow-Up Validation Meeting, San Diego, California, October 9 - 11, 1992

Meredith Powers
Swansea School
Providence, RI

Roger Larsen
Campbell County SD
Gillett, WY

Beth Schieber
Kingfisher Schools
Okarche, OK

Elizabeth Litchfield
Westwood School District
Emmerson, NJ

Larry Barretto
Maplewood Elementary School
Coral Springs, FL

Anne Gregory
Durham Public Schools
Durham, NC

Debra Davidson
Westhampton Beach UFSD
Patchogue, NY

Eugenia Constantinou
Prince George's County School
Silver Spring, MD

Eunice Coakley
Greenville School
Greenville, SC

Nancy Livingston
Brigham Young University
Salt Lake City, UT

Susan McIntyre
University Wisconsin-Eau Claire
Eau Claire, WI

Clyde Colwell
Norfolk Public School
Norfolk, VA

Jo Prather
Mississippi Department of Education
Jackson, MS

Mary Orear
Camden-Rockport HS & MS
Rockport, ME

Shelia Potter
Michigan Department of Education
Lansing, MI

Gene Jongsma
IRA Subcommittee Member
San Antonio, TX

Peggy Dutcher
Michigan Education Assessment Program
Lansing, MI

Martha Carter
Milwaukee Public Schools
Milwaukee, WI

Mark Conley
Michigan State University
Holt, MI

Figure F-6

Meeting Participants, NAEP Reading Revisit
Validation Meeting, Saint Louis, Missouri, October 14 - 16, 1994

Jody Alexander
Madison No. 1
Phoenix, AZ

Evelyn Alford
East Baton Rouge Public Schools
Baton Rouge, LA

Winfrey Bates
Mannsville Elementary School
Mannsville, KY

Joyce Boone
John Strange Elementary School
Indianapolis, IN

Linda Brooks
Alcorn County Public Schools
Corinth, MS

Katie Burnham
Pa Wau Lu Middle School
Gardnerville, NV

Martha Carter
Milwaukee Public Schools
Milwaukee, WI

Carol Case
Mirabean B. Lamar High School
Houston, TX

Molly Chun
Applegate Elementary School
Portland, OR

Roseine Church
Cheyenne, WY

Connie Clayton
Franklin High School
Franklin, WV

David Colburn
Flathead High School
Kalispell, MT

Brenda Creel
Jessup Elementary School
Cheyenne, WY

Pam Diamond
Hellgate Middle School
Missoula, MT

Caroline Downs
Worland Middle School
Worland, WY

Esther Dunnington
Grandview High School
Grandview, MO

Sandra Forsythe
Green Valley High School
Henderson, NV

David Fredette
Westborough High School
Westborough, MA

Cynthia Freeman
Maryville High School
Maryville, TN

Rita Gallagher
Roswell, NM

Lorraine Gerhart
Elmbrook Middle School
Elm Grove, WI

Maria Valeri-Gold
Georgia State University
Atlanta, GA

Figure F-6 (continued)

Meeting Participants, NAEP Reading Revisit
Validation Meeting, Saint Louis, Missouri, October 14 - 16, 1994

Bill Hammond
GA Department of Education
Atlanta, GA

Sally Hellman
Las Vegas, NV

Grace Herr
West Linn High School
West Linn, OR

Sarah Herz
Coleytown Middle School
Westport, CT

Susan Hodgin
Moscow Public Schools
Moscow, ID

Beverly Hoffmaster
Berkeley Heights Elementary School
Martinsburg, WV

Roberta Horton
Custer County High School
Miles City, MT

Lory Johnson
Iowa Department of Education
Des Moines, IA

Ruth Johnson
Holmes High School
Covington, KY

Theresa Lowe
Rancho Viejo School
Yuma, AZ

Ruby Mayes
S.P. Waltrip High School
Houston, TX

Robert McKean
Havre Public Schools
Havre, MT

Pamela McNair
Lemon G. Hine Jr. High School
Washington, DC

Daniel McQuagge
Delta State University
Cleveland, MS

Cheryl Miller
Buchanan Elementary School
Baton Rouge, LA

Donna Miller
Chinook High School
Chinook, MT

Lynn Minderman
Honeoye Falls-Lima Public Schools
Honeoye Falls, NY

John Morrissey
Huntley Project Elementary School
Worden, MT

Pamela Perryman
Selah Middle School
Selah, WA

Kathleen Sanders
Los Angeles Unified S.D.
Wilmington, CA

Helen Schotanus
NH Department of Education
Concord, NH

Terrence Smith
Verona School
Battle Creek, MI

Figure F-6 (continued)

Meeting Participants, NAEP Reading Revisit
Validation Meeting, Saint Louis, Missouri, October 14 - 16, 1994

Faith Stevens
Haslett Public Schools
Haslett, MI

Richard Telfer
Univ. of Wisconsin-Whitewater
Whitewater, WI

Cara Terry
Lakewood High School
St. Petersburg, FL

James Thompson
Simpson-Waverly School
Hartford, CT

Patsy Turner
Great River Co-operative
West Helena, AR

Toni Walters
Oakland University
Rochester, MI

Barbara Watson
Agricola Elementary School
Lucedale, MS

Florence Wakuya
Hawaii Department of Education
Honolulu, HI

Janet Williams
Bluewell Elementary School
Bluefield, WV

Sarah Williams
Maryville Middle School
Maryville, TN

Philip Yeaton
Concord, NH

Katie Young
Louisiana Department of Education
Baton Rouge, LA

Appendix G

SETTING THE NAEP ACHIEVEMENT LEVELS FOR THE 1994 U.S. HISTORY AND GEOGRAPHY ASSESSMENTS

Appendix G

SETTING THE NAEP ACHIEVEMENT LEVELS FOR THE 1994 U.S. HISTORY AND GEOGRAPHY ASSESSMENTS

Mary Lyn Bourque
National Assessment Governing Board

INTRODUCTION

Since 1984, NAEP has reported the performance of students in the nation and for specific subpopulations on a 0-to-500 proficiency scale. The history and development of the scales in U.S. history and geography are described in Chapter 11 of this report.

The 1988 NAEP legislation[1] created an independent board, the National Assessment Governing Board (NAGB), responsible for setting policy for the NAEP program. The 1994 NAEP reauthorization[2] continued many of the Board's statutory responsibilities, including "developing appropriate student performance standards for each age and grade in each subject area to be tested under the National Assessment." Consistent with this directive, and striving to achieve one of the primary mandates of the statute "to improve the form and use of NAEP results," the Board has been developing student performance standards (called achievement levels by NAGB) on the National Assessment since 1990.

Setting achievement levels is a method for setting standards on the NAEP assessment that identify what students should know and be able to do at various points along the proficiency scale. The initial policy definitions and the preliminary descriptions of the achievement levels were presented to panelists along with an illustrative framework for more in-depth development and refinement of the levels. Panelists were asked to craft the recommended descriptions of the three levels from the specific framework developed for the NAEP assessment with respect to the content and skills to be assessed. The operationalized descriptions were refined throughout the level-setting process, and validated with a supplementary group of judges subsequent to the level-setting meetings. Panelists were also asked to select illustrative exercises associated with each level, selecting from the released exercises in the NAEP item pool those sample items that best exemplified the full range of performance of the intervals between levels. The emphasis in operationalizing the definitions and in identifying and selecting exemplar items and papers was to represent the full range of performance from the lower level to the next higher level. The details of the implementation procedures are outlined in the remainder of this appendix.

---

[1] Public Law 100-297. (1988). National Assessment of Educational Progress improvement act (Article No. USC 1221). Washington, DC.

[2] Public Law 103-382. (1994). Improving America's schools act. Washington, DC.

PREPARATION FOR THE U.S. HISTORY AND GEOGRAPHY LEVEL-SETTING MEETINGS

It is important for the planning of any standard-setting effort to know how various process elements interact with each other. For example, panelists interact with premeeting materials, meeting materials (i.e., the assessment items, rating forms, rater feedback, and so forth), each other, and the project staff. All of these elements combine to promote or degrade what has been called intrajudge consistency and interjudge consensus (Friedman & Ho, 1990).

Previous research has conceptualized the effects of two major kinds of interaction: (1) people interacting with text (Smith & Smith, 1988), and (2) people interacting with each other (Curry, 1987; Fitzpatrick, 1989). To assess the effects of textual and social interaction and adjust the standard-setting procedures accordingly, a pilot study in each content area was conducted as the first phase of the 1994 initiative.

Further, the National Academy of Education (NAE) had recommended in its final policy report of the 1992 procedures that methodology changes were needed in the level-setting process (NAE, 1993). Before applying these recommendations in the operational aspects of level-setting, American College Testing (ACT), the contractor responsible for implementing the design, piloted a holistic approach recommended by the NAE.

Finally, four procedures for rating polytomous exercises were piloted (ACT, 1995a; ACT, 1995b). Only one was used in the operational level-setting based on pilot study results. The first method, *Proportional* method, asked panelists to estimate the proportion of examinees at the Basic level who would respond correctly to the item at each score point in the polytomous score scale; the estimates were repeated for the Proficient and Advanced levels. Four-point score scales required 12 estimates per item; three-point score scales required nine estimates per item. The second method, *Modified Proportional* method, required panelists to estimate the proportion of examinees at the Basic level who would respond correctly to the item with a score greater than or equal to 2 (partial correct); the estimates were repeated for the Proficient and Advanced levels. Here, each item required three estimates per item. The third method, *Hybrid* method, asked panelists to select from a sample of student responses three papers that represented the performance at the borderline of the Basic, Proficient, and Advanced levels (nine papers in all). The fourth method, *Mean Score Estimation* method, asked panelists to estimate the mean on the score scale, for borderline Basic, Proficient, and Advanced levels (three estimates in total).

The holistic approach recommended by the NAE was implemented on a block-by-block basis. That is, panelists were asked to estimate the score for borderline Basic, Proficient, and Advanced examinees for the first block as a whole upon completion of their ratings; this was repeated for two blocks overall, three blocks overall, etc. through all the blocks presented. At the end of all ratings, panelists had estimated, in total score points correct, the expected performance of examinees at the borderline of each level.

RESULTS OF THE 1994 PILOT STUDIES

Three of the four procedures for rating test items were implemented in the geography pilot (Proportion, Hybrid, and Mean Estimation), and two of the most promising procedures along with the fourth procedure were piloted in the U.S. history pilot (Proportion, Hybrid, and Modified Proportion). The methods piloted varied substantially in the time required to complete, the complexity of the procedure, the simplicity and user-friendliness of the method, and the resulting cut scores.

The Proportional method consistently resulted in the highest cut scores for polytomous exercises. It required a short time to implement, but did not meet the simplicity and user-friendliness criteria. The Mean method tended to result in the lowest cut scores, and could be implemented in a reasonable period of time. The Hybrid method yielded results similar to the Mean method, however, this method required nearly six hours for the first round of ratings. Finally, the Modified Proportional method resulted in unusually low cut scores, low enough to be suspect. The final recommendation of the contractor was to use the Mean method, but extending the panelists' training to include a paper-selection procedure for all exercises in the rating pool (similar to the Hybrid method). Further details on the various methods and the detailed analyses can be found in the contractor's report of the pilot studies (ACT, 1995b; ACT, 1995c).

Table G-1
Comparison of Pilot Study Cut Scores (SD) in U.S. History and Geography
(Cut Scores on ACT NAEP-Like Scale)

Method	Subject	Grade/Group	Basic	Proficient	Advanced
Proportional	U.S. History	4/A	155.0 (3.7)	164.5 (7.3)	174.4 (7.1)
	U.S. History	8/B	146.9 (6.4)	167.2 (6.4)	177.8 (4.9)
	Geography	4/A	152.5 (6.1)	169.2 (12.5)	177.8 (12.3)
	Geography	8/B	154.6 (8.9)	170.9 (5.6)	182.6 (7.8)
Hybrid	U.S. History	4/B	148.9 (3.9)	155.9 (2.1)	172.8 (5.3)
	U.S. History	12/A	148.2 (5.0)	162.5 (5.2)	172.9 (6.8)
	Geography	4/B	150.9 (4.7)	164.9 (2.9)	176.7 (4.8)
	Geography	12/A	146.8 (7.6)	163.0 (3.5)	178.9 (4.6)
Modified Proportional	U.S. History	8/A	137.7 (6.7)	159.1 (5.4)	173.3 (3.5)
	U.S. History	12/B	130.8 (1.2)	158.3 (6.7)	174.5 (6.6)
Mean Score Estimation	Geography	8/A	148.9 (7.3)	162.3 (5.1)	175.5 (5.0)
	Geography	12/B	159.4 (8.0)	171.0 (6.1)	185.3 (6.7)

1994 LEVEL-SETTING PANELS

Eighty-nine panelists representing the four NAEP regions were selected from the 221 nominees and invited to participate in the geography level-setting process. In U.S. history 77 panelists participated, selected from a nominee pool of 219 persons. They represented teachers at grades 4, 8, and 12, nonteacher educators, and members of the noneducator (general public) community. The group was balanced by

BEST COPY AVAILABLE

gender, race/ethnicity, NAEP regions of the country, community type (i.e., low SES or not low SES), district size, and school type (i.e., public or nonpublic).

1994 PROCESS FOR DEVELOPING THE ACHIEVEMENT LEVELS

Panelists selected for each subject area were convened on separate occasions for a five-day level-setting process. The same agenda was followed for both subjects. In the opening sessions, panelists were provided 'advance organizers' to help them see the complete picture of what they would be doing for the remaining days. An overview, *via* a computerized presentation, demonstrated each step in the process, the reasons for each step, and the interconnections between them. Each panelist was given a 'briefing booklet' that described each task to be performed during each session, purpose of the task, and how to perform the task.

During the first two days panelists were given a brief overview of NAEP and NAGB, a presentation on the policy definitions of the achievement levels, a review of the NAEP assessment frameworks, and a summary of the factors that influence item difficulty. The purpose of the presentations was to focus the panelists' attention on the assessment framework, and to emphasize the fact that panelists' work was directly related to the NAEP assessment, not to the subject matter domain as a whole. In addition, all panelists completed and self-scored an appropriate grade-level form of the NAEP assessment. The purpose of this exercise was to familiarize panelists with the test content and scoring protocols—as well as time constraints—before beginning the formal training for the level-setting activities.

The policy definitions are as follows:

Basic	This level represents partial mastery of prerequisite knowledge and skills that are fundamental for proficient work at each grade.
Proficient	This level represents solid academic performance for each grade assessed. Students reaching this level have demonstrated competency over challenging subject matter, including subject-matter knowledge, application of such knowledge to real world situations, and analytical skills appropriate to the subject matter.
Advanced	This higher level signifies superior performance.

Panelists received training in the frameworks and the preliminary achievement levels descriptions through a series of exercises designed to provide them experience in working with the preliminary descriptions as operationalized during framework development. These descriptions reflect what students *should know and be able to do* at each level. In addition, panelists were expected to become familiar with the various exercise formats, scoring guides, and in the case of polytomous exercises, the scoring rubrics. They were also given the opportunity to review student responses to selected constructed-response exercises so that they could begin to crystallize their conception of borderline performance for each level. These were important as prior activities to the item rating process. Specific training in each task was provided in a general session to ensure standardization in instructions. Grade-level facilitators reinforced the large-group training sessions and answered questions for panelists in grade-level groups. Tasks were then completed as designed. This sequence was followed for all tasks in the five-day session.

Following training in the Angoff method for dichotomously-scored items and the Mean method for polytomous-scored items, the judges began the three-round rating process. In Round 1, and all subsequent rounds, panelists rated about one-half the total number of exercises in the grade-level pool. Following each round, panelists' ratings were key-entered and analyzed to produce feedback information during the remaining rounds of ratings. Following the various rounds, participants were given item difficulty data for all items in their rating pool, interjudge and intrajudge consistency information, and examples of student responses at or near their estimated cut scores. These types of data provided panelists with a 'reality check' against which to compare their ratings. They could then adjust their ratings in subsequent rounds if they thought an adjustment was necessary.

Before Round 3, panelists' ratings were reanalyzed and additional information was presented to panelists concerning intrajudge variability. For each panelist, the intrajudge variability information consisted of those items that they had rated differently than items having similar difficulty, taking into consideration the panelist's aggregated item ratings. That is, the panelists' aggregated item ratings were converted to the theta (θ) scale. All items rated by the panelists were then analyzed as to the panelist's achievement level (θ) in comparison to actual student performance on the items. The observed item rating from each panelist was contrasted with an expected item rating. Those items with the largest differences between observed and expected ratings were identified. Panelists were given this information and asked to review each of these items and decide if their Round 2 ratings still accurately reflected their best judgments of the items. The intrajudge consistency data was to be used to flag items for reconsideration in the final round of rating. Tables G-2 and G-3 display the cut scores for each subject area on the ACT NAEP-like scale, as well the 'percent correct data' across the grade-level item pool.

Following Round 3 panelists were given 'consequences data', that is, panelists were given close approximations of the percentages of students who would score at or above each achievement level based on the cut scores that had been set during the rating process. They were asked to comment on the data specific to their grade. In general, panelists' expectations were not unlike the data being presented to them. The number of panelists who would have opted to change the levels was very small, and stood as confirmation of the process and their understanding of it.

ADDITIONAL ANALYSES OF THE 1994 DATA

Additional analyses were completed to examine the effects of item type, panelist type, and rating group membership on the item ratings. Significant differences in cut scores were found for dichotomous and polytomous exercises in 17 out of the 18 cut scores examined. The only exception was in geography, grade 4 Basic. Table G-4 shows the differences between polytomous and dichotomous cut scores for Round 3 (since these were the cut scores adopted). In all but the exception, the cut scores for the polytomous items are higher. Further, the higher the grade the larger the difference, and the differences are generally larger for U.S. history than for geography.

Each grade-level group was divided in half with each group rating about half the item in the grade-level pool. The purpose of this assignment to half groups was to allow for variance estimation and thus, estimates of reliability. The effect of rating group membership was significant depending on grade level, rating round, or subject area. The significant differences in Round 3 (the round that was finally adopted) included geography, grade 4 Proficient and Advanced, and grade 12 all levels. In U.S. history, there were no significant differences.

BEST COPY AVAILABLE

Table G-2

Geography Cut Scores (Standard Deviations) by Grade and Round,
*and Percent Correct Data**

Grade/Round	Basic	Proficient	Advanced
4/Round 1	152.8 (9.1)	168.1 (3.9)	179.1 (3.7)
Round 2	151.9 (7.4)	167.8 (3.9)	178.8 (3.8)
Round 3	151.9 (6.9)	167.8 (3.9)	178.8 (3.7)
% Correct	34.7%	61.4%	80.8%
8/Round 1	146.9 (5.1)	163.1 (4.8)	174.2 (6.5)
Round 2	150.0 (5.1)	164.4 (4.3)	176.1 (5.3)
Round 3	150.4 (5.4)	164.9 (4.5)	176.9 (5.2)
% Correct	37.6%	60.5%	80.0%
12/Round 1	149.3 (8.6)	164.1 (7.3)	177.0 (6.9)
Round 2	151.4 (5.0)	166.2 (5.5)	180.9 (4.8)
Round 3	150.8 (4.5)	166.0 (5.2)	180.8 (4.4)
% Correct	39.4%	64.4%	83.6%

****Note***: *Percent correct data are estimates of the percentage of possible points required for a score at the lower borderline of each achievement level. Read: "Students would have to get at least 83.6% of the possible points on the items to score at the Advanced level in Grade 12."*

Table G-3

U.S. History Cut Scores (Standard Deviations) by Grade and Round, and Percent Correct Data

Grade/Round	Basic	Proficient	Advanced
4/Round 1	157.1 (7.3)	170.1 (3.8)	180.6 (5.1)
Round 2	155.1 (4.6)	168.4 (3.4)	177.8 (4.9)
Round 3	155.5 (4.8)	168.8 (3.4)	178.3 (4.6)
% Correct	33.1%	56.7%	75.2%
8/Round 1	155.2 (6.8)	171.5 (4.1)	185.1 (6.8)
Round 2	154.2 (5.1)	171.1 (3.9)	184.5 (5.7)
Round 3	154.3 (5.5)	171.2 (3.7)	184.6 (6.2)
% Correct	37.1%	61.6%	80.7%
12/Round 1	159.3 (9.0)	170.7 (3.1)	182.5 (5.2)
Round 2	159.6 (6.2)	171.3 (2.5)	182.9 (4.4)
Round 3	159.6 (4.9)	171.6 (2.6)	1833.6(4.0)
% Correct	41.5%	63.2%	82.3%

***Note**: *Percent correct data are estimates of the percentage of possible points required for a score at the lower borderline of each achievement level. Read: "Students would have to get at least 82.3% of the possible points on the items to score at the Advanced level in Grade 12."*

Table G-4

Mean Differences Between Polytomous and Dichotomous Cut Scores

Subject/Grade	Basic	Proficient	Advanced
Geography/Grade 4	6.258	3.682	4.970
Geography/Grade 8	8.918	6.762	8.596
Geography/Grade 12	9.646	8.582	9.142
U.S. History/Grade 4	10.094	2.884	3.416
U.S. History/Grade 8	16.758	8.190	4.718
U.S. History/Grade 12	16.982	7.462	7.798

BEST COPY AVAILABLE

The effect of panelist type on cut scores was nonsignificant. Since there was no grade-by-panelist interaction found, the data were pooled across grades. Tables G-5 displays the cut scores by panelist type for Round 3 (the cut scores finally adopted).

Table G-5
Mean Cut Scores (Standard Deviation) by Panelist Type

Subject/Panelist Type	Panelist Type	Basic	Proficient	Advanced
Geography/Grade 4	Teacher (N=16)	150.9 (5.0)	167.9 (3.6)	179.1 (2.8)
	Educator (N=5)	150.5 (11.9)	168.6 (4.2)	180.0 (2.9)
	Noneducator (N=7)	152.0 (7.6)	166.8 (4.8)	178.0 (6.0)
Geography/Grade 8	Teacher (N=17)	149.9 (5.0)	165.0 (4.0)	176.6 (4.8)
	Educator (N=3)	147.2 (3.9)	162.3 (5.5)	176.7 (9.6)
	Noneducator (N=8)	151.4 (6.8)	166.1 (5.3)	179.1 (4.2)
Geography/Grade 12	Teacher (N=18)	150.4 (4.9)	165.8 (5.1)	179.8 (4.4)
	Educator (N=3)	150.3 (5.2)	168.9 (5.4)	183.1 (3.4)
	Noneducator (N=8)	151.9 (3.5)	165.5 (5.8)	181.4 (4.4)
U.S. History/Grade 4	Teacher (N=15)	154.6 (4.2)	167.7 (3.5)	177.5 (3.9)
	Educator (N=4)	156.9 (2.8)	169.4 (1.7)	179.6 (1.8)
	Noneducator (N=9)	155.1 (5.9)	169.0 (3.9)	179.1 (7.0)
U.S. History/Grade 8	Teacher (N=15)	155.5 (5.0)	171.1 (4.3)	184.9 (6.3)
	Educator (N=5)	154.8 (5.7)	170.7 (4.2)	185.4 (7.1)
	Noneducator (N=9)	152.6 (5.0)	171.6 (3.1)	185.1 (4.6)
U.S. History/Grade 12	Teacher (N=14)	160.6 (5.2)	171.6 (2.5)	183.4 (3.4)
	Educator (N=4)	157.6 (4.6)	171.9 (2.7)	183.3 (4.8)
	Noneducator (N=9)	160.4 (4.8)	171.9 (2.9)	183.9 (5.0)

ACHIEVEMENT LEVELS DESCRIPTIONS

In addition to an in-depth study of the preliminary descriptions prior to the rating process, panelists were given the opportunity to review the descriptions and make modifications before each round of ratings. The suggested modifications were then submitted to the content staff (persons who had been involved with the development of the subject area framework and/or the development of exercises in that content area). A joint decision by the content staff determined whether such modifications represented significant substantive changes, in which case, they were incorporated into the next iteration of the descriptions.

In the final analysis, very few changes were recommended in either subject area. The complete and approved descriptions are included at the end of this appendix.

SELECTING EXEMPLAR ITEMS

On the final day of the achievement level-setting process, panelists reviewed all items from the item pools in U.S. history or geography. Although all items would not be released, there were some anticipated limitations with the proposed released item pool. So as a safety net, all items were examined by the panelists as possible items illustrative of the achievement levels for their grade.

Exercises are organized in blocks consisting of several items, usually employing each of the three item formats, (i.e., multiple-choice, short constructed-response, and extended constructed-response). Before the review process, all exercises were categorized using statistical criteria recommended by the Technical Advisory Committee on Standard Setting (TACSS), the group that advises ACT on technical decisions throughout the process. Items having an average rp=.50 for scores within the achievement level range were included in the list of items submitted to panelists for their consideration. Items were listed at the lowest level for which they met the criterion. Constructed-response items were treated as unique items at each score point, and thus, could meet the criterion n-1 times, where n = the number of score categories. Items could be recommended as exemplars at a higher level than the statistical criterion placed them (based on content), but could not be placed at a lower level, since the rp would fall below .50 at a lower level.

In the final analysis, it was decided that a complete block of items would be released for purposes of representing the content of the assessment, while student responses to selected constructed-response items would best represent the achievement levels. Both a complete block and selected student responses appear in the content area reports that have been released.

MAPPING THE LEVELS ONTO THE NAEP SCALE

The process of mapping panelists' ratings to the NAEP scales used *item response theory* (IRT). IRT provided statistically sophisticated methods for determining the expected performance of examinees on particular test items in terms of an appropriate measurement scale. The same measurement scale simultaneously described the characteristics of the test items and the performance of the examinees. Once the item characteristics were set, it was possible to determine precisely how examinees were likely to perform on the test items at different points of the measurement scale.

The panelists' ratings of the NAEP test items were likewise linked, by definition, to the expected performance of examinees at the theoretical achievement level cut points. It was therefore feasible to use the IRT item characteristics to calculate the values on the measurement scale corresponding to each achievement level. This was done by averaging the item ratings over panelists for each achievement level and then simply using the item characteristics to find the corresponding achievement level cut points on the IRT measurement scale. This process was repeated for each of the NAEP U.S. history and geography scales within each grade (4, 8, and 12).

For the multiple-choice and short constructed-response items that were dichotomously scored, the judges each rated half of the items in the NAEP pool in terms of the expected probability that a student at a borderline achievement level would answer the item correctly, based on the judges' operationalization of the policy definitions and the factors that influence item difficulty. To assist the judges in generating consistently scaled ratings, the rating process was repeated twice, with feedback. Information on consistency among different judges and on the difficulty of each item[3] was fed back into the first repetition (Round 2), while information on consistency within each judge's set of ratings was fed back into the second repetition (Round 3). The third round of ratings permitted the judges to discuss their ratings among themselves to resolve problematic ratings. The mean final rating of the judges aggregated across multiple-choice and short constructed-response items yielded the threshold values for these items in the percent correct metric. These cut scores were then mapped onto the NAEP scale (which is defined and scored using item response theory, rather than percent correct).

For extended constructed-response items, judges were asked to estimate the mean score on the rating score scale for the borderline performance at each achievement level. The overall panelists' mean was mapped onto the NAEP scale in a manner similar to that used for the items scored dichotomously.

The final cut score for each achievement level was a weighted average of the cut score for the multiple-choice and short constructed-response items and the cut score for the extended constructed-response items, with the weights being proportional to the information supplied by the two classes of items.

In the final stage of the mapping process, the achievement level cut points on the IRT measurement scale were combined over content areas and rescaled to the NAEP score scale. Weighted averages of the achievement level cut points were computed. The weighting constants accounted for the measurement precision of the test items evaluated by the panelists, the proportion of items belonging to each NAEP content area, and the linear NAEP scale transformations. These weighted averages produced the final cut points for the basic, proficient, and advanced achievement levels within each grade.

1994 RESEARCH STUDIES ON THE ACHIEVEMENT LEVELS

Two studies were completed to examine the various aspects of the validity of the NAEP achievement levels. The first, *Similarity Classification Study (SCS),* was designed to compare the classifications of students according to the achievement levels descriptions by students' teachers with the classification of the same students according to their performance on a specially designed version of NAEP that yielded individual scores. The second, *Booklet Classification Study (BCS),* was designed to compare the performances of students (as demonstrated in their NAEP booklet) with the knowledge and skills described in the NAEP achievement levels descriptions.

The SCS was conducted with only grade 8 students and their teachers who participated in the either the ALS pilot or the ALS meeting. Thirty-five teachers (19 from geography and 16 from U.S. history) as well as 1,791 students (956 from geography and 835 from U.S. history) participated in the study. Each student was administered a special form of the NAEP (4 blocks) requiring 100 minutes of testing time. The special form was developed by ETS to meet certain minimum criteria, was administered by Westat under the same conditions as a standardized NAEP administration, and scored by NCS using identical scoring procedures to the 1994 NAEP administration. The purpose of the longer-length NAEP was to be able to directly estimate NAEP scores for students without having to use the conditioning model and plausible values technology.

---

[3] Item difficulty estimates were based on a preliminary, partial set of responses to the national assessment.

Students' performances in the appropriate subject area were classified by their teachers based on their knowledge of the students relative to the assessment framework and the achievement levels descriptions. Students' actual performances were subsequently classified according to their scores on the extended NAEP assessment. The results of these two classifications were compared.

The details of the analyses conducted by ACT can be found in the report of these studies (ACT, 1995a). However, in general, students' performance on the assessment was lower than their teachers' judgment of the students' performance. Overall, teachers classified their students in higher categories of the achievement levels than did the results of student performance on the assessment. This would suggest, of course, that the achievement levels are set higher than is warranted by student performance.

The second study, BCS, was conducted with panelists at grades 4, 8, and 12, and in both subject areas, geography and U.S. history. Sixty panelists in total, representing both subjects and all three grades, participated in this study. The 10 panelists in each grade/subject group examined 40 student booklets, chosen to be representative of the complete assessment and the framework. The only exception to this was grades 4 and 12 in U.S. history, where only 38 booklets were available. Prior to the panel meeting, all booklets in the study had been classified according to the achievement levels cut scores as either Basic, Proficient, or Advanced. Panelists were asked to classify the booklets in the same way, but without having the knowledge of the empirical classification provided through the plausible values for each booklet.

The details of this study may also be found in the ACT (1995a) report. As in the earlier study the consistency of results between the two classifications were mixed. In general, however, the booklet classifications by panelists were either on target or lower than the empirical classifications. This result would suggest that the achievement levels cut scores were set lower than would be warranted by student performances.

As in any study of this nature, the results are subject to many caveats, and ACT points out several of these in their report. These studies certainly are not intended to be definitive of the validity of the achievement levels. They are an indication, however, that additional data analyses need to be completed in order to probe more fully the technical characteristics of the levels as adopted by the NAGB.

Figure G-1

Final Descriptions of 1994 U.S. History Achievement Levels

GRADE 4 *BASIC* (195)	Fourth-grade students performing at the Basic level should be able to identify and describe a few of the most familiar people, places, events, ideas, and documents in American history. They should be able to explain reasons for celebrating most national holidays, have some familiarity with the geography of their own state and the United States, and be able to express in writing a few ideas about a familiar theme in American history.
PROFICIENT (243)	Fourth-grade students performing at the Proficient level should be able to identify, describe and comment on the significance of many historical people, places, ideas, events, and documents. They should interpret information from a variety of sources, including texts, maps, pictures, and timelines. They should be able to construct a simple timeline from data. These students should recognize the role of invention and technological change in history. They should also recognize the ways in which geographic and environmental factors have influenced life and work.
ADVANCED (276)	Fourth-grade students performing at the Advanced level should have a beginning understanding of the relationships between people, places, ideas, events, and documents. They should know where to look for information, including reference books, maps, local museums, interviews with family and neighbors, and other sources. They should be able to use historical themes to organize and interpret historical topics, and to incorporate insights from beyond the classroom into their understanding of history. These students should understand and explain the role of invention and technological change in history. They should also understand and explain the ways in which geographic and environmental factors have influenced life and work.
GRADE 8 *BASIC* (252)	Eighth-grades students performing at the Basic level should be able to identify and place in context a range of historical people, places, events, ideas, and documents. They should be able to distinguish between primary and secondary sources. They should have a beginning understanding of the diversity of the American people and the ways in which people from a wide variety of national and cultural heritages have become part of a single nation. Eighth-grade students at the Basic level should also have a beginning understanding of the fundamental political ideas and institutions of American life and their historical origins. They should be able to explain the significance of some major historical events.
PROFICIENT (294)	Eighth-grade students performing at the Proficient level should be able to explain the significance of people, places, events, ideas, and documents, and to recognize the connection between people and events within historical contexts. They should understand and be able to explain the opportunities, perspectives and challenges associated with a diverse cultural population. They should incorporate geographic, technological, and other considerations in their understanding of events and should have knowledge of significant political idea and institutions. They should be able

Figure G-1 (continued)
Final Descriptions of 1994 U.S. History Achievement Levels

to communicate ideas about historical themes while citing evidence from primary and secondary sources to support their conclusions.

ADVANCED
(327)

Eighth-grade students performing at the Advanced level should recognize significant themes and movements in history and begin to understand particular events in light of these themes and movements. They should have an awareness of continuity and change over time and be able to draw relevant analogies between past events and present-day situations. They should be able to frame questions about historical topics and use multiple sources to develop historical generalizations and interpretations. They should be able to explain the importance of historical themes, including some awareness of their political, social, and economic dimensions.

GRADE 12
BASIC
(294)

Twelfth-grade students performing at the Basic level should be able to identify the significance of many people, places, events, dates, ideas, and documents in U.S. history. They should also recognize the importance of unity and diversity in the social and cultural history of the United States, and an awareness of America's changing relationships with the rest of the world. They should have a sense of continuity and change in history and be able to relate relevant experience from the past to their understanding of contemporary issues. They should recognize that history is subject to interpretation and should understand the role of evidence in making an historical argument.

PROFICIENT
(294)

Twelfth-grade students performing at the Proficient level should understand particular people, places, events, ideas, and documents in historical context, with some awareness of the political, economic, geographic, social, religious, technological, and ideological factors that shape historical settings. They should be able to communicate reasoned interpretations of past events, using historical evidence effectively to support their positions. Their written arguments should reflect some in-depth grasp of issues and refer to both primary and secondary sources.

ADVANCED
(355)

Twelfth-grade students achieving at the Advanced level should demonstrate a comprehensive understanding of events and sources of U.S. history. Recognizing that history is subject to interpretation, they should be able to evaluate historical claims critically in light of the evidence. They should understand that important issues and themes have been addressed differently at different times and that America's political, social, and cultural traditions have changed over time. They should be able to write well-reasoned arguments on complex historical topics and draw upon a wide range of sources to inform their conclusions.

Figure G-2

Final Descriptions of 1994 Geography Achievement Levels

GRADE 4 *BASIC* (187)	Students should be able to use words or diagrams to define basic geographic vocabulary; identify personal behaviors and perspectives related to the environment and describe some environmental and cultural issues in their community; use visual and technological tools to access information; identify major geographic features on maps and globes; be able to read and draw simple maps, map keys and legends; demonstrate how people depend upon, use, and adapt to the environment; and give examples of the movement of people, goods, services, and ideas from one place to another. In addition to demonstrating an understanding of how individuals are alike and different, they should demonstrate a knowledge of the ways people depend on each other.
PROFICIENT (240)	Students should be able to use fundamental geographic knowledge and vocabulary to identify basic geographic patterns and processes; describe an environmental or cultural issue from more than one perspective; and read and interpret information from visual and technological tools such as photographs, maps, and globes, aerial photography, and satellite images. They should be able to use number and letter grids to plot specific locations; understand relative location terms; and sketch simple maps and describe and/or draw landscapes they have observed or studied. Proficient students should be able to illustrate how people depend upon, adapt, to, and modify the environment, describe and/or illustrate geographic aspects of a region using fundamental geographic vocabulary and give reasons for current human migration; discuss the impact a location has upon cultural similarities and differences; and be able to demonstrate how an event in one location can have an impact upon another location.
ADVANCED (276)	Students should be able to use basic knowledge and vocabulary to describe global patterns and processes; describe ways individuals can protect and enhance environmental quality; describe how modifications to the environment may have a variety of consequences; explain differing perspectives that apply to local environmental or cultural issues; and demonstrate an understanding of forces that result in migration, changing demographics, and boundary changes. They should be able to solve simple problems by applying information learned through working with visual and technological tools such as aerial and other photographs, maps and globes, atlases, news media, and computers. They should be able to construct models and sketch and label maps of their own state, the United States, and the world; use them to describe and compare differences, similarities, and patterns of change in landscapes; and be able to predict the impact a change in one location can have on another. They should be able to analyze the ways individuals and groups interact.

Figure G-2 (continued)
Final Descriptions of 1994 Geography Achievement Levels

GRADE 8 *BASIC* (242)	Students should possess fundamental knowledge and vocabulary of concepts relating to patterns, relationships, distance, direction, scale, boundary, site, and situation; solve fundamental locational questions using latitude and longitude; interpret simple map scales; identify continents and their physical features, oceans, and various countries and cities; respond accurately to descriptive questions using information obtained by use of visual and technological tools such as geographic models and/or translate that information into words; explain differences between maps and globes; and find a wide range of information using environments, and provide evidence showing how physical habitat can influence human activity. They should be able to define a region and identify its distinguishing characteristics. Finally, they should be able to demonstrate how the interaction that takes place between and among regions is related to the movement of people, goods, services, and ideas.
PROFICIENT (282)	Students should possess a fundamental geographic vocabulary; understand geography's analytical concepts; solve locational questions requiring integration of information from two or more sources, such as atlases or globes; compare information presented at different scales; identify a wide variety of physical and cultural features and describe regional patterns. Students should be able to respond accurately to interpretive questions using geography's visual and technological tools and translate that information into patterns; identify differences in map projections and select proper projections for various purposes; and develop a case study working with geography's analytical concepts. In addition, students should be able to describe the physical and cultural characteristics of places; explain how places change due to human activity; explain and illustrate how the concept of regions can be used as a strategy for organizing and understanding Earth's surface. Students should be able to analyze and interpret data bases and case studies as well as use information from maps to describe the role that regions play in influencing trade and migration patterns and cultural and political interaction.
ADVANCED (315)	Students should have a command of extensive geographic knowledge, analytical concepts, and vocabulary; be able to analyze spatial phenomena using a variety of sources with information presented at a variety of scales and show relationships between them; and use case studies for spatial analysis and to develop maps and other graphics. Students should be able to identify patterns of climate, vegetation, and population across Earth's surface and interpret relationships between and among these patterns, and use one category of a map or aerial photograph to predict other features of a place such as vegetation based on climate or population density based on topographic features. Students should also be able to relate the concept of region to specific places and explain how regions change over time due to a variety of factors. They should be able to profile a region of their own design using geographic concepts, tools, and skills.

Figure G-2 (continued)
Final Descriptions of 1994 Geography Achievement Levels

GRADE 12 *BASIC* (270)	Students should possess a knowledge of concepts and terms commonly used in physical and human geography as well as skills enabling them to employ applicable units of measurement and scale when solving simple locational problems using maps and globes. They should be able to read maps; provide examples of plains, plateaus, hills, and mountains; and locate continents, major bodies of water, and selected countries and cities. They should be able to interpret geographic data and use visual and technological tools such as charts, tables, cartograms, and graphs; know the nature of and be able to identify several basic types of map projections; understand the basic physical structure of the planet; explain and apply concepts such as continental drift and plate tectonics; and describe geography's analytical concepts using case studies. Students should have a comprehensive understanding of spatial relationships including the ability to recognize patterns that exist across Earth in terms of phenomena, including climate regions, time zones, population distributions, availability of resources, vegetation zones, and transportation and communication networks. They should be able to develop data bases about specific places and provide a simple analysis about their importance.
PROFICIENT (305)	Students should have an extensive understanding and knowledge of the concepts and terminology of physical and human geography. They should be able to use geographic concepts to analyze spatial phenomena and to discuss economic, political, and social factors that define and interpret space. They should be able to do this through the interpretation of maps and other visual and technological tools, through the analysis of case studies, the utilization of data bases, and the selection of appropriate research materials. Students should be able to design their own maps based on descriptive data; describe the physical and cultural attributes of major world regions; relate the spatial distribution of population to economic and environmental factors; report both historical and contemporary events within a geographic framework using tools such as special purpose maps, and primary and secondary source materials.
ADVANCED (339)	Students should possess a comprehensive understanding of geographic knowledge and concepts; apply this knowledge to case studies; formulate hypotheses and test geographic models that demonstrate complex relationships between physical and human phenomena; apply a wide range of map skills; develop maps using fundamental cartographic principles including translating narratives about places and events into graphic representations, and use other visual and technological tools to perform locational analysis and interpret spatial relationships. Students should also be able to undertake sophisticated analysis from aerial photographs or satellite imagery and other visuals. Advanced students should be able to develop criteria assessing issues relating to human spatial organization and environmental stability and, through research skills and the application of critical thinking strategies, identify alternative solutions. They should be able to compile data bases from disparate pieces of information and from these data develop generalizations and speculations about outcomes when data change.

Figure G-3

Participants, NAEP U.S. History Pilot Study Meeting
St. Louis, Missouri, August 11 - 15, 1994

Barbara J. Allen
Horace Mann School
Denver, CO 80211

Mary M. Allen
The Town of Newton
Newton, NH 03858

Rick Alvarez
Lattie Coor School
Avondale, AZ 85323

Chris Becker
Stenner Glen School
San Luis Obispo, CA 93405

Polly Burkhart
Jewett Academy
Winter Haven, FL 33881

Judy Card
Gust Elementary
Denver, CO 80219

Mark Charlton
Indian Trails Middle School
Winter Springs, FL 32708

Eartha Clarke
Somerville Board of Education
Somerville, NJ 08876

Russell Clayton
Pine Mountain Middle School
Kennesaw, GA 30144

Janice Conley
St. Albans, WV 25177

Robert D. Cross
University of Virginia
Charlottesville, VA 22903

Becky Davenport
Winter Springs Elementary
Winter Springs, FL 32708

Sandi Feinn
North Hillsborough School
Hillsborough, CA 94010

Deena Kaye Music Fisher
Woodward High School
Woodward, OK 73801

Anne Elizabeth Fitzsimmons
LaCima Middle School
Tucson, AZ 85704

James W. Fleet
John F. Kennedy High School
Denver, CO 80227

Douglas Frutiger
North Atlanta High School
Atlanta, GA 30305

Diane Hamilton
James Island High School
Charleston, SC 29412

Jean Hamlin
Mesa Public Schools
Mesa, AZ 85218-1096

Margaret Hayes
Intervale School
Parsippany, NJ 07054

Todd Hohensee
Kennesaw Elementary School
Kennesaw, GA 30144

David Hunt
Clarence-Lowden High School
Clarence, IA 52216

Figure G-3 (continued)
Participants, NAEP U.S. History Pilot Study Meeting
St. Louis, Missouri, August 11 - 15, 1994

Edgar Jackson
Robbins, IL 60472

Renee L. Jenson
White Bear Lake Area Public Schools
White Bear Lake, MN 55110

David Jongetjes
Bonesteel Fairfax Schools
Bonesteel, SD 57317

Paul Karlowicz
University High School
Tucson, AZ 85711

Joseph C. Kudless
Raritan Valley Community College
Somerville, NJ 08876

Johnson Lancaster
Progressive Emporium
St. Louis, MO 63130

Kathy Laster
San Luis Obispo Chamber of Commerce
San Luis Obispo, CA 93401

Jane Lawson
Glendale Union High School District
Glendale, AZ 85301

Jerryelyn Leonard-Jones
Curie High School
Chicago, IL 60632

Randy Lyon
Hoover Elementary School
Dubuque, IA 52001

Mary Lee Martin
Board of Education
Albuquerque, NM 87125

Alice M. Mitchell
McGlone Elementary School
Denver, CO 80239

Rev. George Moore
Newton, NH 03858

Sandra R. Mozingo
Horn Lake High School
Horn Lake, MS 38637

Jan Murray
John Ross Elementary School
Edmond, OK 73003

Kathleen Nemann
Roberts Paideia School
Cincinnati, OH 45204

Ben Olk
Breck School
Golden Valley, MN 55422

Pauline W. Pendarvis
City of Port St. Joe
Port St. Joe, FL 32456

Tammy Porter
Grand Avenue Middle School
Milwaukee, WI 53233

Kenneth Reid
Mountain Pointe High School
Phoenix, AZ 85044

R. Kraig Rittenhouse
Warren Central High School
Vicksburg, MS 39180

Lorraine Roberson
Maryland Avenue Elementary
Milwaukee, WI 53211

Figure G-3 (continued)

Participants, NAEP U.S. History Pilot Study Meeting
St. Louis, Missouri, August 11 - 15, 1994

Jerry L. Seimears
Kent, WA 98031

Richard Shetley
Vineyard Junior High School
Alta Loma, CA 91737

David Spector
Wiscasset, ME 04578

Wanda Strassburg
San Luis Sourdough
San Luis Obispo, CA 93401

Karen Talus
Oregon Episcopal School
Portland, OR 97223

Jean Tate
Independence Middle School
Yukon, OK 73099

Nancy Taylor
St. Michael School
Worthington, OH 43085

Anita Thompson
West Lauderdale H.S.
Collinsville, MS 39325

Tom Tinder
West Virginia State Bar
Charleston, WV 25311-2204

Mark Turner
Mt. Zion Elementary School
Jonesboro, GA 30236

Christy Wells
Gwyneth Ham
Yuma, AZ 85365

Kenneth Whitcomb
Crocker Middle School
Hillsborough, CA 94010

Sandy White
Orchard Elementary School
Gahanna, OH 43230

Ronald Wild
Alexander Hamilton High School
Milwaukee, WI 53220

Bernice Wilson
Dunbar Elementary School
East St. Louis, IL 62207

Janet Youel
St. John's Hospital
Maplewood, MN 55109

Table G-4

Participants, NAEP U.S. History Achievement-Level Setting Meeting
St. Louis, Missouri, December 2 - 6, 1994

Sharon Andrews
S & S Co.
Colton, NY 13625

Carl Atkinson
Wisenhickon Middle School
Ambler, PA 19002

Denise Bedard-Comstock
Fairfield Center School
Fairfield, VT 05455

Helen Bibler
Bitter Root Valley Historical Society
Hamilton, MT 59840

Sharon Bixby
Ludlow Insurance Co.
Ludlow, VT 05149

Patrice Brewer
Goshen-Lempster Cooperative
East Lempster, NH 03605

Jackie Brooks
Ligon Middle School
Raleigh, NC 27601

Mitchell Brown
Temecula Valley High School
Temecula, CA 92592

Ren Bryan
Cummings High School
Burlington, NC 27215

Barbara Campbell
Scotland County R-I
Memphis, MO 63555

Ed Cannaday
Stigler Public Schools
Stigler, OK 74462

Linda Clark
St. Thomas Moore School
Brooklyn, OH 44144

Paul Cohen
South Brunswick Board of Education
Monmouth Junction, NJ 08852

James Collins
Quaker Street NY 12141

Cinda Cooper
South Zanesville Intermediate School
South Zanesville, OH 43701

Scott Cooper
Chamber of Commerce
Prineville, OR 97754

May Cornwall
Public Works Department
St. Croix, VI 00820

Clarence Currie
Superior Court of Lake County
Crown Point, IN 46307

Pat Dye
Venice, FL 34292

Phyllis Ehlers
Oakdale High School
Oakdale, LA 71463

Elizabeth Ehrlenbach
Surry Elementary School
Surry, ME 04684

Joe Etheredge
Riverway Mortagee Company
Houston, TX 77024

Table G-4 (continued)
Participants, NAEP U.S. History Achievement-Level Setting Meeting
St. Louis, Missouri, December 2 - 6, 1994

Joyce Fisher
Fritsche Middle School
Milwaukee, WI 53207

Rose Marie Floyd
Jefferson Davis High School
Montgomery, AL 36111

Stephanie Fullerton
Pearl, MS 39208

Angela Garcia-Sims
Coronado, CA 92118

Melvin Garrison
Independence National Historical Park
Philadelphia, PA 19106

Gordon Gillespie
Bowman Museum
Prineville, OR 97754

Syd Golston
Mountainside Middle School
Scottsdale, AZ 85259

Paula Hadgis
Medina Junior High School
Medina, OH 44256

William Harris
Cedarville High School
Cedarville, OH 45314

Peggy Hawse
Moorefield, WV 26836

Peg Hill
Office of Education
San Bernardino, CA 92410

Karen Hoppes
Lakeridge High School
Lake Oswego, OR 97034

Arpad Horvath
Sevastopol Schools Junior High School
Sturgeon Bay, WI 54235

Geraldine Hudley
Dougherty Comprehensive High School
Albany, GA 31705

Gail Huschle
St. Croix Catholic School
Stillwater, MN 55082

Elizabeth Johnson
Athens Drive High School
Raleigh, NC 27606

Janet Keeton
Hickory Hills Elementary School
Marietta, GA 30060

Michael Koury
The Old Army Press
Fort Collins, CO 80525

Cynthia LaFountain
Colton, NY 13625

Irene Lees
Parkway School
Paramus, NJ 07652

Sandra Leighton
Limestone Community School
Limestone, ME 04750

Thelma Longboat
Black Rock Academy #51
Buffalo, NY 14207

Chris Magoc
Alle-Kiski Valley Historical Society
Tarentuh, PA 15084

Table G-4 (continued)

Participants, NAEP U.S. History Achievement-Level Setting Meeting
St. Louis, Missouri, December 2 - 6, 1994

Jill Marett
Wallace Elementary/Middle School
Wallace, SC 29596

Joan McKown
Lucas Local Schools
Lucas, OH 44843

Pat Mitchell
Gwinnett County Board of Education
Lawrenceville, GA 30245

Sheila Moore
Hamlin Park School #74
Buffalo, NY 14208

Terry Moore
Greater Gary Chamber
504 Broadway, Suite 324
Gary, IN 46402

Kristin Palmquist
Department of Education
721 Capitol Mall
Sacramento, CA 95814

Randall Pellow
Shippensburg University
1871 Old Main Drive
Shippensburg, PA 17257

Yann Pirrone
Smithfield Middle School
800 Main Street
Smithfield, VA 23430

Michael Porter
Mary Ann Elementary School
Newark, OH 43055

Joann Prewitt
Office of Instructional Design
Jackson, MS 39201

Sandra Reeves
Clark Junior High School
East St. Louis, IL 62205

Charlie Rhodarmer
Scottish Tartans Museum
Franklin, NC 28734

Judy Rigdon
Millbrook Junior High School
Millbrook, AL 36054

Patricia Roach
Arkansas Tech University
Russellville, AR 72801

Ann Rogers
Texas Education Agency
Austin, TX 78701

Norman Ross
Greater East St. Louis Chamber
East St. Louis, IL 62201

Gerald Schmitz
TMS Service
Menomonee Falls, WI 53051

Paul Schramm
Western Dubuque High School
Epworth, IA 52045

David Scott
Benedum Airport
Bridgeport, WV 26330

Dick Scott
Morrison, CO 80465

Diane Sheets
New Kensington Area Chamber of Commerce
New Kensington, PA 15068

Table G-4 (continued)

Participants, NAEP U.S. History Achievement-Level Setting Meeting
St. Louis, Missouri, December 2 - 6, 1994

Janet Shockey
Shanksville Stonycreek School
Shanksville, PA 15560

Jay Simonsen
Roosevelt School
Racine, WI 53402

Patricia Sirois
Bangor High School
Bangor, ME 04401

Marlene Smoot
Children & Family Services
East St. Louis, IL 62201

Elizabeth Snyder
Eisenhower Middle School
Succasunna, NJ 07874

Rick Steber
Bonanza Publishing Co.
Prineville, OR 97754

Jack Stetson
Fredericktown Intermediate School
Fredericktown, OH 43019

Joe Suminski
Franklin, NC 28734

Cherry Sveen
Macon County Historical Society
Franklin, NC 28734

Ann Toplovich
Tennessee Historical Society
Nashville, TN 37243

Gary VerDouw
Capital Electric Cooperative
Bismarck, ND 58502

Melody Walcott
Federal Public Defender
Christiansted,
St. Croix, VI 00822

Dacia Walsh
Liberty Elementary School
Sapula, OK 74066

David Whitehorse
CA State University at San Marcos
San Marcos, CA 92096

Evona Willhoite
Salina Elementary School
Salina, OK 74365

Georgia Wilson
Lovejoy High School
Lovejoy, GA 30257

Kathryn Woodard
Emerson Elementary School
Tulsa, OK 74106

Pat Yarbrough
Southwest High School
Macon, GA 31206

BEST COPY AVAILABLE

Table G-5

Participants, NAEP Geography Pilot Study Meeting
St. Louis, Missouri, July 14 - 18, 1994

Margaret Dwen Andrews
Kennelly School
Hartford, CT 06114

Betty Arivett
Kirby Elementary School
Kirby, AR 71950

Jerry Aschermann
Missouri Western State College
St. Joseph, MO 64507

Edward F. Bergman
Lehman College
The Bronx, NY 10468-1589

Laurie K. Bricker
Houston, TX 77071

Diann Brown
Patrick Henry Local Schools
Hamler, OH 43524

Linda P. Bryant
State of GA West Unit - Probation
Atlanta, GA 30310

Susan Burcham
Fountain Hills Jr/Sr High School
Fountain Hills, AZ 85268

Russ Bush
Buckhannon-Upshur Middle School
Buckhannon, WV 26201

James Case
Vinson High School
Huntington, WV 25704

William Charlton
Conway Middle School
Orlando, FL 32812

Ann Curtis Cox
Granby Memorial Middle School
Granby, CT 06035

Don Robert Cressall
Cook Elementary School
Syracuse, UT 84075

Barbara Dick
Hills County Council PTA/PTSA
Tampa, FL 33602

Anneke Dodge
Longfellow School
Long Beach, CA 90807

Jerry Dogget
Cassia County School District 151
Burley, ID 83318

Bert Donahue
General Motors Corporation
Buffalo, NY 14240-0021

Mary Beth Dunleavy
Arlington, TX 76017

Mary E. Fehlman
St. Joseph School
Elgin, IL 60120

The Honorable Ronald Goodman
Dickey County Courthouse
Ellendale, ND 58436

Alan Hardway
Ritchie County Middle School
Ellenboro, WV 26346

Russell Hazeltine
Conneaut High School
Conneaut, OH 44030

Table G-5 (continued)

Participants, NAEP Geography Pilot Study Meeting
St. Louis, Missouri, July 14 - 18, 1994

Sarah M. Hunt
McEvoy Middle School
Macon, GA 31206

Donald B. Hutton
Delaware & Hudson Railway Co.
Buffalo, NY 14206

Alison Hyde
NY State School Boards Association
Albany, NY 12210

Ismael Legarreta
El Paso, TX 79915

Cathy Lord
Henry County Public Schools
McDonough, GA 30253

Patrick McNeal
Holdingford Public Schools
Holdingford, MN 56340

Mary Myrick
Dresden Elementary School
Dresden, TN 38225

Joyce Nixon
East Clayton Elementary School
Ellenwood, GA 30049

Bobby Occeana
C.G. White Middle School
Powellsville, NC 27967

Bruce W. Odom
Smith Barney Shearson
El Paso, TX 79901

Michael A. Ostapuk
Estrella Junior High School
Phoenix, AZ 85033

George R. Parish
Robert A. Millikan High School
Long Beach, CA 90815

Bill Perry
Traverse Company
Sutton, MA 01590

Mary Ann Polve
Mesa Public Schools
Mesa, AZ 85205

Sandy Prisloe
EnviroGraphics
Old Saybrook, CT 06475

Norma Jean Remington
Centerville Junior High
Centerville, UT 84014

Lois Sargent
Pupils Service Educational Center
Memphis, TN 38104

Arlene Schwartz
Salem Senior High School
Salem, OH 44460

Kirby Shaw
Ernest Childers Middle School
Broken Arrow, OK 74012

Thomas E. Sherer
Kilderatlas Publishing Company
Old Lyme, CT 06371

James Donald Smith
Fort Worth, TX 76109-1611

Dan R. Sotomayor
Vista Del Sol Conference Center
El Paso, TX 79935

Table G-5 (continued)

Participants, NAEP Geography Pilot Study Meeting
St. Louis, Missouri, July 14 - 18, 1994

Donald E. Steiss
A.J. Nystrom Company
Tampa, FL 33629

William T. Strack
Leukemia Society of America
Tampa, FL 33607

Debbie Turney
Hulbert Heat and Air
Hulbert, OK 74441

George Welsh
Tortolita Middle School
Tucson, AZ 85741

Jean Wendt
Bloomington, IL 61704

Doris White
Clear Creek Trading Post
Hulbert, OK 74441

Joe White
Normal High School
Normal, IL 61762

Becky Widener
University of Missouri-Columbia
Columbia, MO 65211

Patricia Yates
Saints Joseph and John School
Strongsville, OH 44136

Vicki Young
Central Davis Junior High School
Layton, UT 84041

Joan Zuckerman
Ridgeview Middle School
Gaithersburg, MD 20878

Table G-6

Participants, NAEP Geography Achievement-Level Setting Meeting
St. Louis, Missouri, November 12 - 16, 1994

Nancy Abel
Greenbriar West Elementary School
Fairfax, VA 22033

Marie Andel
North Glendale Elementary School
Kirkwood, MO 63122

Rowletta Anderson
Clayton Middle School
Clayton, OK 74536

Diane Baker
The Map Center
Huntsville, AL 35802

Tom Ballenger
Bishop Kelley High School
Tulsa, OK 74135

Katherine Barrett
Georgetown East School
Annapolis, MD 21403

Frederick Birnie
Department of Transportation
Deland, FL 32720

Joyce Blanton
Robert E. Lee School
Tullahoma, TN 37388

Janice Boehme
City of Hillsboro
Hillsboro, WI 54634

Fredia Booker
South Atlanta High School
Atlanta, GA 30354

Carlton Brinkley
Enderly Heights Elementary School
Buena Vista, VA 24416

LaVeta Chase
Healdton Middle School
Healdton, OK 73438

Frederic Cheney
School District Eleven
Colorado Springs, CO 80903

David Cole
Univ. of Northern Colorado
Greeley, CO 80639

Jill Cook
Wabasha-Kellogg Elementary School
Wabasha, MN 55981

Lois Crowley
Coralville Central Elementary School
Coralville, IA 52241

Alice Dewees
Foster Middle School
Tulsa, OK 74129

Betty Dixon
Hilldale Middle School
Muskogee, OK 74403

Jeanne Doremus
Vineland Board of Education
Vineland, NJ 08360

Sue Dougherty
Dougherty Trucking Company
Plain City, OH 43064

Corwin Drake
Montclair High School
Montclair, NJ 07042

R. Michael Eckenroth
Classical High School
Providence, RI 02903

Figure G-6 (continued)
Participants, NAEP Geography Achievement-Level Setting Meeting
St. Louis, Missouri, November 12 - 16, 1994

Michael Edelstein
Marlboro Elementary School
Marlboro, VT 05344

Karla Escalera
McNair Middle School
San Antonio, TX 78252

Nick Fielder
Environment & Conservation ?????
Nashville, TN 37211

Laurie Finucane
Cypress Elementary School
Fontana, CA 92335

Jana Flores
UCLA History-Geography Project
Los Angeles, CA 90024

Thomas Franke
Gunderson Clinic-Hillsboro
Hillsboro, WI 54634

Dana Glei
Child Trends, Inc.
Washington, DC 20008

Willie Graham
Colonial Williamsburg Foundation
Petersburg, VA 23803

Nancy Gray
Dairy Farm
Groveton, NH 03582

James Hantula
Price Laboratory School
Cedar Falls, IA 50613

Doris Harrison
First State Bank
Gainesville, TX 76240

Shirley Herrick
Mason High School
Mason, MI 48854

Tim Heydt
Thompson School District
Loveland, CO 80538

Matt Hodge
Winman Junior High School
Warwick, RI 02886

June Hollis
Brandon High School
Brandon, MS 39042

Judith Horchler
Huron High School
Huron, OH 44839

Lynne Hortenstine
Hortenstine Farms
Brookfield, MO 64628

Reid Hughes
Daytona Beach, FL 32115

Arlyne Kaus
Browns Valley, MN 56219

Marianne Kenney
CO State Department of Education
Denver, CO 80203

Eleanor Kenyon
Oakton High School
Vienna, VA 22181

Eva Kinsey
Harper High School
Atlanta, GA 30331

Figure G-6 (continued)

Participants, NAEP Geography Achievement-Level Setting Meeting
St. Louis, Missouri, November 12 - 16, 1994

Robert Lee
Green River College
Auburn, WA 98002

Barbara Loverich
Hobart High School
Hobart, IN 46342

V. Colleen Mahaney
Winship Magnet School
Macon, GA 31204

Joan Maier
Univ. of Houston-Clear Lake
Houston, TX 77058

Barbara Malm
Blue Hill Consolidated School
Blue Hill, ME 04614

Lyn Malone
Barrington Middle School
Barrington, RI 02806

Susan Markle
Our Lady of Perpetual Help School
Toledo, OH 43614

Fred Matteucci
Shorecrest Senior High School
Seattle, WA 98155

Sandra McQueen
Sutton Middle School
Atlanta, GA 30327

Curt Miller
Oakman High School
Oakman, AL 35579

Patricia Miller
Conley Hills Elementary School
East Point, GA 30344

Judy Miller
Park View High School
Sterling, VA 22064

Reginald Moore
Daytona Beach, FL 32114

Chris Moseley
Superior Industries, Inc.
Ormond Beach, FL 32174

Justin Mutrux
Raymondville R-VII School
Raymondville, MO 65555

Nora Obney
Lake Local Middle School
Hartville, OH 44632

Rawhide Papritz
Green River Community College
Auburn, WA 98002

Ann Pelchat
Groveton, NH 03582

Sherel Perry
Auburn High School
Auburn, AL 36830

William Pogue
Puxico, MO 63960

Sharon Popowski
Holy Rosary Elementary School
North Mankato, MN 56003

Joseph Rach
Fort Worth, TX 76107

Ken Rathje
Village of Downers Grove
Downers Grove, IL 60515

Figure G-6 (continued)

Participants, NAEP Geography Achievement-Level Setting Meeting
St. Louis, Missouri, November 12 - 16, 1994

Linda Reynolds
Groveton, NH 03582

Mary Rowland
Idalou Independent School District
Idalou, TX 79329

Joan Schmidt
Gallipolis, OH 45631

Maureen Spaight
E.R. Martin Junior High
East Providence, RI 02914

Daniel Sparler
The Northwest School
Seattle, WA 98121

Michelle Stasa
Linden McKinley High School
Columbus, OH 43211

Merle Temple
Bell South Co.
Augusta, GA 30901

Sherry Terrill
Holtville Middle School
Holtville, CA 92250

Ellen Thompson
St. Luke School
Barrington, RI 02806

Mahala Tising
Sylacauga High School
Sylacauga, AL 35150

Zarema Turner
Mt. Hebron School
Montclair, NJ 07043

Betty Upchurch
Mendoza Elementary School
Mesa, AZ 85205

Kurt Waters
Annandale High School
Annandale, VA 22003

Karen Wilcox
Finley Elementary School
Holtville, CA 92250

Margaret Wilkinson
Carson Junior High School
Mesa, AZ 85201

Pat Williams
Downers Grove National Bank
Downers Grove, IL 60515

Judith Wiseman
George Rogers Clark High School
Winchester, KY 40391

Donna Zeglarski
Roselle Park Middle School
Roselle Park, NJ 07204

Figure G-7

Participants, NAEP U.S. History and Geography
Similarity Classification Study, Kansas City, MO, March 4 - 6, 1995

Rowletta Anderson
Clayton Middle School
Clayton, OK 74536

Patrice Brewer
Goshen-Lempster Cooperative
East Lempster, NH 03605

Jackie Brooks
Ligon Middle School
Raleigh, NC 27601

Diann Brown
Patrick Henry Local Schools
Hamler, OH 43524

Mark Charlton
Indian Trails Middle School
Winter Springs, FL 32708

William Charlton
Conway Middle School
Orlando, FL 32812

Linda Clark
St. Thomas Moore School
Brooklyn, OH 44144

Ann Curtis Cox
Granby Memorial Middle School
Granby, CT 06035

Betty Dixon
Hilldale Middle School
Muskogee, OK 74403

Anne Fitzsimmons
LaCima Middle School
Tucson, AZ 85704

Paula Hadgis
Medina Junior High School
Medina, OH 44256

James Hantula
Price Laboratory School
Cedar Falls, IA 50613

Alan Hardway
Ritchie County Middle School
Ellenboro, WV 26346

Matt Hodge
Winman Junior High School
Warwick, RI 02886

Arpad Horvath
Sevastopol Schools-Junior High
Sturgeon Bay, WI 54235

Sarah M. Hunt
McEvoy Middle School
Macon, GA 31206

Gail Huschle
St. Croix Catholic School
Stillwater, MN 55082

Lyn Malone
Barrington Middle School
Barrington, RI 02806

Joan McKown
Lucas Local Schools
Lucas, OH 44843

Sandra McQueen
Sutton Middle School
Atlanta, GA 30327

Sheila Moore
Hamlin Park School #74
Buffalo, NY 14208

Michael A. Ostapuk
Estrella Junior High School
Phoenix, AZ 85033

Figure G-7 (continued)

Participants, NAEP U.S. History and Geography
Similarity Classification Study, Kansas City, MO, March 4 - 6, 1995

Norma Jean Remington
Centerville Junior High
Centerville, UT 84014

Judy Rigdon
Millbrook Junior High School
Millbrook, AL 36054

Richard Shetley
Vineyard Junior High School
Alta Loma, CA 91737

Elizabeth Snyder
Eisenhower Middle School
Succasunna, NJ 07874

Maureen Spaight
E.R. Martin Junior High
East Providence, RI 02914

Jean Tate
Independence Middle School
Yukon, OK 73099

Sherry Terrill
Holtville Middle School
Holtville, CA 92250

Anita Thompson
West Lauderdale High School
Collinsville, MS 39325

Zarema Turner
Mt. Hebron School
Montclair, NJ 07043

George Welsh
Tortolita Middle School
Tucson, AZ 85741

Kenneth Whitcomb
Crocker Middle School
Hillsborough, CA 94010

Vicki Young
Central Davis Junior High School
Layton, UT 84041

Donna Zeglarski
Roselle Park Middle School
Roselle Park, NJ 07204

Figure G-8

Participants, NAEP U.S. History and Geography Booklet Classification Study, St. Louis, MO, April 8 - 11, 1995

James Adams
Sevastopol High School
Sturgeon Bay, WI 54235

Linda Adum
Tampa, FL 33647

Mitchell Aho
Londonderry Middle School
Londonderry, NH 03053

Fred Alexander
Nantahala Power & Light Company
Franklin, NC 28734

Robert Allekotte
Philadelphia, PA 19134

Warren Baker
Stevenson Elementary School
Mesa, AZ 85208

Gaylynn Becker
Department of Public Instruction
Bismarck, ND 58505

Sam Bell
Lake Forest High School
Lake Forest, IL 60045

Diane Bernhardt
Thomas Jefferson Middle School
Valpraiso, IN 46383

John Brierley
Hermosa Beach, CA 90254

Marshall Brown
Alton Middle School
Alton, IL 62002

John Carnesecca
Temecula Valley High School
Temecula, CA 92592

Kate Collins
St. Monica School
Dallas, TX 75229

Mary Jo Costello
Englewood, CO 80110

Elenor Cox
College Station, TX 77840

John Craig
Illinois State Board of Education
Springfield, IL 62704

William A. DeGrazia
Bethel Middle School
Bethel, CT 06801

Betty DeVore
Ontario Middle School
Mansfield, OH 44906

Mary Dobbs
Dover High School
Dover, AR 72837

Judy Dollard
Eisenhower Middle School
Kansas City, KS 66109

Patricia Ferguson
116 Luther Brown Road
Colerain, NC 27924

Doug Finch
Kelly Elementary School
Moore, OK 73160

Karin Finks
C.W. McGraw
Yuma, AZ 85364

BEST COPY AVAILABLE

Figure G-8 (continued)

Participants, NAEP U.S. History and Geography
Booklet Classification Study, St. Louis, MO, April 8 - 11, 1995

Rita Haelman
Eisenhower Middle School
Rockford, IL 61107

Robert Harber
Fredric Remington High School
Whitewater, KS 67154

Gay Harris
East St. Louis, IL 62201

Walt Herscher
Appleton School District
Appleton, WI 54913

Joyce Hoover
American Council on Education
Washington, DC 20036

Ann Hopkins
Minneapolis, MN 55410

Georgeanne Hribar
Virginia Beach, VA 23464

Ann Humble
George Rogers Clark School
Winchester, KY 40391

Jon Iverson
Watertown High School
Watertown, SD 57201

Richard Janson
Janson Industries
Canton, OH 44706

Carol Jones
Fredericktown, OH 43019

Caryl Jones
Ventana Vista School
Tucson, AZ 85715

Junko Kajita
San Diego, CA 92129

Geraldo Kaprosy
Harding High School
St. Paul, MN 55106

Bill Kellogg
Bow, NH 03304

Mary Knightly
Kensington Elementary School
Kensington, NH 03833

Albert Larson
Chicago, IL 60607

Marshall Levy
Milton, MA 02186

Carol Lewallen
St. Nicholas School
Chattanooga, TN 37421

Minoo Monakes
Coker College
Hartsville, SC 79550

James Nason
Perry & Morrill, Inc.
Bangor, ME 04401

Kristin Nyquist
Amherst, NY 14226

Judy Rapp
Pierre Junior High School
Pierre, SD 57501

Susan Scott
Ritchie County Clerk
Harrisville, WV 26362

Figure G-8 (continued)

Participants, NAEP U.S. History and Geography
Booklet Classification Study, St. Louis, MO, April 8 - 11, 1995

Sheila Spellacy
East Hartford Middle School
East Hartford, CT 06108

Marcia Spencer
Stevens High School
Claremont, NH 03743

John Stedman
Manchester High School
Manchester, CT 06040

Susan Temple
Department of Public Instruction
Raleigh, NC 27601-2825

Sondra Towne
Indianapolis Public School #106
Indianapolis, IN 46218

Nancy Turner
Mesa Education Association
Mesa, AZ 85203

Karen Tyrrell
Los Angeles, CA 90005

Betty Vollmer
The Vollmer Farm
Bunn, NC 27508

Sheila Washington
City of Scottsboro
Scottsboro, AL 35768

Julie Weiss
Department of Education
Malden, MA 02148

Linda Whitney
Montebello, CA 90640

John Willmer
State University College
Cortland, NY 13045

Jeanne Witzig
BRW, Inc.
Minneapolis, MN 55415

Appendix H

CONSTRUCTED-RESPONSE ITEM SCORE STATISTICS

Appendix H

CONSTRUCTED-RESPONSE ITEM SCORE STATISTICS

This appendix contains information about the constructed-response items included in the scaling of data from the 1994 main assessments of reading, U.S. history, and geography.

The information in the tables includes, for each subject area and age/grade, the NAEP item numbers for each of the constructed-response items included in scaling, and the block that contains the item. The tables also indicate the codes from the NAEP database that denote the range of responses and the correct responses. A portion of the responses to the constructed-response items were scored twice for the purpose of examining rater reliability. For each item, the number of papers with responses that were scored a second time is listed, along with the percent agreement between raters and an index of reliability based on those responses. Cohen's Kappa (Cohen, 1968) is the reliability estimate used for dichotomized items. For items that are not dichotomized (i.e., polytomous items), the intraclass correlation coefficient is used as the index of reliability.

Table H-1

*Score Range, Percent Agreement, and Cohen's Kappa**
for the Dichotomously Scored Constructed-Response Reading Items
Used in Main Assessment Scaling, Age 9/Grade 4

Item	Block	Range of Response Codes	Correct Response Codes	Sample Size	Percent Agreement	Cohen's Kappa
R012201		1 - 2	2	584	91	0.83
R012208		1 - 2	2	482	95	0.91
R012206		1 - 2	2	509	95	0.92
R012210		1 - 2	2	433	95	0.88
R012104		1 - 2	2	561	93	0.88
R012106		1 - 2	2	534	93	0.88
R012102		1 - 2	2	589	94	0.89
R012108		1 - 2	2	530	95	0.89
R012112		1 - 2	2	361	90	0.77
R012109		1 - 2	2	497	97	0.94
R012706		1 - 2	2	527	92	0.82
R012705		1 - 2	2	518	95	0.88
R012710		1 - 2	2	355	93	0.87
R012702		1 - 2	2	580	95	0.89
R012703		1 - 2	2	536	96	0.92
R012611		1 - 2	2	436	95	0.91
R012604		1 - 2	2	569	95	0.90
R012601		1 - 2	2	587	93	0.85
R015802		1 - 2	2	585	91	0.81
R012002		1 - 2	2	557	95	0.92
R012008		1 - 2	2	517	92	0.86
R012004		1 - 2	2	561	92	0.86
R012010		1 - 2	2	460	92	0.85
R012504		1 - 2	2	586	96	0.94
R012508		1 - 2	2	552	97	0.95
R012506		1 - 2	2	521	94	0.91
R012503		1 - 2	2	584	88	0.79
R012511		1 - 2	2	492	93	0.87

****Note:*** *Cohen's Kappa is a measure of reliability that is appropriate for items that are dichotomized.*

Table H-2

Score Range, Percent Agreement, and Intraclass Correlation for the Polytomously Scored Constructed-Response Reading Items Used in Main Assessment Scaling, Age 9/Grade 4

Item	Block	Range of Response Codes	Sample Size	Percent Agreement	Intraclass Correlation
R015703		1 - 3	575	87	0.86
R015709		1 - 3	415	92	0.93
R015704		1 - 3	568	84	0.89
R015702		1 - 3	587	83	0.81
R015705		1 - 3	542	91	0.96
R015806		1 - 3	487	80	0.85
R015803		1 - 3	572	83	0.81
R015807		1 - 3	468	83	0.88
R015809		1 - 3	396	81	0.83
R012204		1 - 4	576	79	0.91
R012111		1 - 4	411	92	0.96
R015707		1 - 4	330	89	0.90
R012708		1 - 4	475	84	0.88
R012607		1 - 4	510	91	0.94
R015804		1 - 4	569	83	0.89
R012006		1 - 4	551	86	0.93
R012512		1 - 4	458	79	0.92

BEST COPY AVAILABLE

Table H-3
*Score Range, Percent Agreement, and Cohen's Kappa**
for the Dichotomously Scored Constructed-Response Reading Items
Used in Main Assessment Scaling, Age 13/Grade 8

Item	Block	Range of Response Codes	Correct Response Codes	Sample Size	Percent Agreement	Cohen's Kappa
R015901		1 - 2		602	93	0.87
R016001		1 - 2		601	89	0.79
R012706		1 - 2	2	606	87	0.76
R012713		1 - 2	2	473	99	0.98
R012705		1 - 2	2	606	90	0.82
R012710		1 - 2	2	542	92	0.85
R012702		1 - 2	2	597	98	0.94
R012703		1 - 2	2	594	88	0.78
R013108		1 - 2	2	566	100	0.99
R013111		1 - 2	2	453	89	0.65
R013110		1 - 2	2	495	89	0.81
R013104		1 - 2	2	601	94	0.88
R013102		1 - 2	2	607	83	0.70
R013105		1 - 2	2	587	92	0.83
R013211		1 - 2	2	521	90	0.80
R013205		1 - 2	2	603	98	0.94
R013207		1 - 2	2	595	93	0.85
R013203		1 - 2	2	610	98	0.90
R013209		1 - 2	2	584	97	0.92
R012611		1 - 2	2	561	94	0.85
R012604		1 - 2	2	605	94	0.89
R012601		1 - 2	2	628	89	0.79
R013008		1 - 2	2	527	93	0.87
R013007		1 - 2	2	567	98	0.96
R013003		1 - 2	2	569	100	0.99
R013001		1 - 2	2	591	96	0.89
R013009		1 - 2	2	542	96	0.87
R013010		1 - 2	2	515	97	0.93
R013005		1 - 2	2	571	95	0.89
R013011		1 - 2	2	490	90	0.82
R013412		1 - 2	2	472	89	0.70
R013409		1 - 2	2	556	92	0.85
R013405		1 - 2	2	602	90	0.82
R013411		1 - 2	2	493	93	0.86
R013402		1 - 2	2	602	99	0.97
R013407		1 - 2	2	549	96	0.91

***Note**: *Cohen's Kappa is a measure of reliability that is appropriate for items that are dichotomized.*

Table H-4

Score Range, Percent Agreement, and Intraclass Correlation for the Polytomously Scored Constructed-Response Reading Items Used in Main Assessment Scaling, Age 13/Grade 8

Item	Block	Range of Response Codes	Sample Size	Percent Agreement	Intraclass Correlation
R016109		1 - 3	547	93	0.91
R016108		1 - 3	593	83	0.86
R016101		1 - 3	593	87	0.91
R016104		1 - 3	578	88	0.86
R016107		1 - 3	574	94	0.98
R016201		1 - 3	593	95	0.85
R016207		1 - 3	596	93	0.96
R016211		1 - 3	590	87	0.83
R016205		1 - 3	588	90	0.90
R016212		1 - 3	558	92	0.94
R016213		1 - 3	556	86	0.83
R016202		1 - 3	603	89	0.79
R015907		1 - 3	507	88	0.89
R015905		1 - 3	577	88	0.91
R015904		1 - 3	587	89	0.83
R015908		1 - 3	398	82	0.83
R015902		1 - 3	599	91	0.93
R016006		1 - 3	567	89	0.91
R016004		1 - 3	582	84	0.86
R016003		1 - 3	585	85	0.88
R016210		1 - 4	580	76	0.89
R016204		1 - 4	597	89	0.89
R012708		1 - 4	586	81	0.87
R013106		1 - 4	591	80	0.89
R015906		1 - 4	544	85	0.87
R013201		1 - 4	617	85	0.91
R013212		1 - 4	526	90	0.93
R016009		1 - 4	539	83	0.92
R016002		1 - 4	588	86	0.90
R012607		1 - 4	589	84	0.90
R013004		1 - 4	582	84	0.93
R013406		1 - 4	596	85	0.94
R013403		1 - 4	607	97	0.99

Table H-5

*Score Range, Percent Agreement, and Cohen's Kappa**
for the Dichotomously Scored Constructed-Response Reading Items
Used in Main Assessment Scaling, Age 17/Grade 12

Item	Block	Range of Response Codes	Correct Response Codes	Sample Size	Percent Agreement	Cohen's Kappa
R013108		1 - 2	2	489	100	1.00
R013111		1 - 2	2	418	85	0.64
R013110		1 - 2	2	437	89	0.82
R013104		1 - 2	2	504	93	0.82
R013102		1 - 2	2	513	85	0.66
R013105		1 - 2	2	486	90	0.81
R013704		1 - 2	2	519	92	0.85
R013701		1 - 2	2	506	91	0.82
R013702		1 - 2	2	513	84	0.68
R013710		1 - 2	2	481	91	0.78
R013708		1 - 2	2	505	86	0.74
R013712		1 - 2	2	404	84	0.69
R015509		1 - 2	2	498	88	0.80
R015505		1 - 2	2	499	91	0.84
R015512		1 - 2	2	485	90	0.81
R015503		1 - 2	2	506	95	0.79
R013211		1 - 2	2	465	89	0.79
R013205		1 - 2	2	498	98	0.88
R013207		1 - 2	2	514	93	0.80
R013203		1 - 2	2	503	98	0.88
R013209		1 - 2	2	491	98	0.96
R013501		1 - 2	2	485	92	0.85
R013505		1 - 2	2	484	89	0.75
R013509		1 - 2	2	333	89	0.76
R013508		1 - 2	2	344	90	0.79
R013503		1 - 2	2	491	94	0.90
R013910		1 - 2	2	527	95	0.92
R013902		1 - 2	2	523	93	0.84
R013904		1 - 2	2	513	98	0.94
R013903		1 - 2	2	515	95	0.91
R013908		1 - 2	2	541	88	0.79
R013913		1 - 2	2	511	94	0.91
R013906		1 - 2	2	536	92	0.87
R013412		1 - 2	2	451	87	0.67
R013409		1 - 2	2	499	96	0.91
R013405		1 - 2	2	515	93	0.83
R013411		1 - 2	2	444	95	0.85
R013402		1 - 2	2	515	97	0.94
R013407		1 - 2	2	510	96	0.90

***Note**: *Cohen's Kappa is a measure of reliability that is appropriate for items that are dichotomized.*

Table H-6

Score Range, Percent Agreement, and Intraclass Correlation for the Polytomously Scored Constructed-Response Reading Items Used in Main Assessment Scaling, Age 17/Grade 12

Item	Block	Range of Response Codes	Sample Size	Percent Agreement	Intraclass Correlation
R016109		1 - 3	496	93	0.92
R016108		1 - 3	493	81	0.85
R016101		1 - 3	507	91	0.93
R016104		1 - 3	523	89	0.84
R016107		1 - 3	532	96	0.98
R016303		1 - 3	495	91	0.94
R016305		1 - 3	464	84	0.89
R016302		1 - 3	509	90	0.89
R016301		1 - 3	514	86	0.87
R016306		1 - 3	427	82	0.88
R016307		1 - 3	393	90	0.93
R016601		1 - 3	512	84	0.85
R016603		1 - 3	502	87	0.80
R016602		1 - 3	513	84	0.88
R016604		1 - 3	500	86	0.86
R016501		1 - 3	512	92	0.88
R016605		1 - 3	499	87	0.82
R016502		1 - 3	512	88	0.88
R016407		1 - 3	425	78	0.83
R016402		1 - 3	510	82	0.83
R016403		1 - 3	491	87	0.90
R016405		1 - 3	502	86	0.87
R016401		1 - 3	508	87	0.87
R016404		1 - 3	500	87	0.87
R016308		1 - 4	395	80	0.87
R013106		1 - 4	509	77	0.89
R013706		1 - 4	508	83	0.85
R015507		1 - 4	515	88	0.94
R015514		1 - 4	505	79	0.90
R013201		1 - 4	507	85	0.92
R013212		1 - 4	470	91	0.94
R016701		1 - 4	491	87	0.89
R016408		1 - 4	428	89	0.92
R013506		1 - 4	456	85	0.91
R013915		1 - 4	149	89	0.93
R013406		1 - 4	516	91	0.97
R013403		1 - 4	504	98	0.99

Table H-7

*Score Range, Percent Agreement, and Cohen's Kappa**
for the Dichotomously Scored Constructed-Response U.S. History Items
Used in Main Assessment Scaling, Age 9/Grade 4

Item	Block	Range of Response Codes	Correct Response Codes	Sample Size	Percent Agreement	Cohen's Kappa
H034501		1 - 2	2	566	100	0.99

Table H-8

*Score Range, Percent Agreement, and Cohen's Kappa**
for the Dichotomously Scored Constructed-Response U.S. History Items
Used in Main Assessment Scaling, Age 13/Grade 8

Item	Block	Range of Response Codes	Correct Response Codes	Sample Size	Percent Agreement	Cohen's Kappa
H034501		1 - 2	2	570	100	0.99

Table H-9

*Score Range, Percent Agreement, and Cohen's Kappa**
for the Dichotomously Scored Constructed-Response U.S. History Items
Used in Main Assessment Scaling, Age 17/Grade 12

Item	Block	Range of Response Codes	Correct Response Codes	Sample Size	Percent Agreement	Cohen's Kappa
H046001		1 - 2	2	524	99	0.99

****Note:*** *Cohen's Kappa is a measure of reliability that is appropriate for items that are dichotomized.*

Table H-10

Score Range, Percent Agreement, and Intraclass Correlation for the Polytomously Scored Constructed-Response U.S. History Items Used in Main Assessment Scaling, Age 9/Grade 4

Item	Block	Range of Response Codes	Sample Size	Percent Agreement	Intraclass Correlation
H033301		1 - 3	532	98	0.972
H031801		1 - 3	596	90	0.945
H031802		1 - 3	585	97	0.986
H032601		1 - 3	569	85	0.911
H030101		1 - 3	513	86	0.922
H033001		1 - 3	567	91	0.930
H034702		1 - 3	576	98	0.972
H028701		1 - 3	581	88	0.945
H028702		1 - 3	571	91	0.954
H031402		1 - 3	521	80	0.893
H035101		1 - 3	515	91	0.952
H029002		1 - 3	439	88	0.939
H029702		1 - 3	557	90	0.933
H029701		1 - 3	572	92	0.929
H030603		1 - 3	577	92	0.956
H030604		1 - 3	575	94	0.957
H033201		1 - 3	542	92	0.949
H028801		1 - 3	561	87	0.905
H035001		1 - 3	542	89	0.918
H032503		1 - 3	388	90	0.905
H034401		1 - 3	575	97	0.988
H030302		1 - 3	409	91	0.789
H032301		1 - 3	560	86	0.898
H031101		1 - 3	560	86	0.902
H028201		1 - 4	601	92	0.940
H032701		1 - 4	565	94	0.976
H031701		1 - 4	595	86	0.924
H030701		1 - 4	592	93	0.965
H029501		1 - 4	564	85	0.952
H030201		1 - 4	473	88	0.899
H034101		1 - 4	585	94	0.933

Table H-11

Score Range, Percent Agreement, and Intraclass Correlation for the Polytomously Scored Constructed-Response U.S. History Items Used in Main Assessment Scaling, Age 13/Grade 8

Item	Block	Range of Response Codes	Sample Size	Percent Agreement	Intraclass Correlation
H037203		1 - 3	555	89	0.842
H037601		1 - 3	582	98	0.980
H033301		1 - 3	533	96	0.979
H038702		1 - 3	568	85	0.833
H038301		1 - 3	566	90	0.949
H041301		1 - 3	566	88	0.826
H042902		1 - 3	573	91	0.936
H032601		1 - 3	536	84	0.898
H041801		1 - 3	550	94	0.941
H041802		1 - 3	534	88	0.703
H039001		1 - 3	510	93	0.935
H042201		1 - 3	561	97	0.938
H033001		1 - 3	549	92	0.948
H034702		1 - 3	566	93	0.921
H035101		1 - 3	569	92	0.925
H035101		1 - 3	569	92	0.925
H043201		1 - 3	586	90	0.933
H043401		1 - 3	582	89	0.936
H043601		1 - 3	575	86	0.870
H043701		1 - 3	583	91	0.943
H043705		1 - 3	577	81	0.833
H040001		1 - 3	564	94	0.959
H033201		1 - 3	528	92	0.949
H038601		1 - 3	562	90	0.850
H035901		1 - 3	567	89	0.901
H035902		1 - 3	580	84	0.849
H039401		1 - 3	572	90	0.904
H035001		1 - 3	558	81	0.913
H035801		1 - 3	561	92	0.876
H036101		1 - 3	574	87	0.926
H034401		1 - 3	572	95	0.962
H037101		1 - 3	577	89	0.906
H043101		1 - 3	498	87	0.877
H040103		1 - 3	555	89	0.930
H043001		1 - 3	523	91	0.863
H053901		1 - 3	552	96	0.965
H040201		1 - 3	554	93	0.949
H038103		1 - 4	583	93	0.899
H042801		1 - 4	583	86	0.870
H041401		1 - 4	566	91	0.956

Table H-11 (continued)
Score Range, Percent Agreement, and Intraclass Correlation for the Polytomously Scored Constructed-Response U.S. History Items Used in Main Assessment Scaling, Age 13/Grade 8

Item	Block	Range of Response Codes	Sample Size	Percent Agreement	Intraclass Correlation
H032701		1 - 4	535	85	0.929
H036402		1 - 4	570	82	0.901
H044001		1 - 4	569	83	0.895
H043501		1 - 4	583	92	0.967
H037401		1 - 4	563	88	0.940
H039901		1 - 4	565	89	0.944
H034101		1 - 4	566	85	0.856

Table H-12

Score Range, Percent Agreement, and Intraclass Correlation for the Polytomously Scored Constructed-Response U.S. History Items Used in Main Assessment Scaling, Age 17/Grade 12

Item	Block	Range of Response Codes	Sample Size	Percent Agreement	Intraclass Correlation
H049503		1 - 3	511	92	0.934
H041301		1 - 3	530	86	0.867
H042902		1 - 3	483	87	0.912
H044702		1 - 3	535	87	0.861
H045301		1 - 3	508	96	0.976
H041801		1 - 3	534	90	0.907
H041802		1 - 3	465	83	0.840
H053304		1 - 3	546	87	0.913
H053201		1 - 3	547	84	0.852
H052902		1 - 3	538	78	0.772
H053103		1 - 3	529	93	0.953
H053102		1 - 3	551	87	0.906
H053003		1 - 3	553	84	0.818
H047601		1 - 3	548	86	0.885
H046301		1 - 3	544	90	0.914
H052601		1 - 3	513	85	0.862
H042201		1 - 3	539	96	0.931
H048301		1 - 3	500	89	0.933
H046101		1 - 3	531	95	0.961
H047501		1 - 3	548	89	0.934
H048901		1 - 3	522	99	0.995
H045102		1 - 3	508	92	0.958
H051102		1 - 3	519	80	0.889
H051101		1 - 3	537	90	0.932
H052701		1 - 3	438	88	0.933
H052501		1 - 3	526	93	0.967
H047401		1 - 3	542	91	0.930
H048202		1 - 3	515	86	0.918
H043101		1 - 3	478	91	0.897
H051002		1 - 3	515	98	0.984
H043001		1 - 3	492	87	0.894
H049701		1 - 3	519	98	0.981
H051301		1 - 3	548	83	0.867
H042801		1 - 4	515	88	0.922
H041401		1 - 4	537	91	0.951
H049601		1 - 4	490	88	0.939
H044301		1 - 4	531	87	0.925
H050201		1 - 4	537	82	0.900
H053601		1 - 4	493	92	0.939
H053501		1 - 4	518	91	0.958

Table H-12 (continued)

Score Range, Percent Agreement, and Intraclass Correlation for the Polytomously Scored Constructed-Response U.S. History Items Used in Main Assessment Scaling, Age 17/Grade 12

Item	Block	Range of Response Codes	Sample Size	Percent Agreement	Intraclass Correlation
H052901		1 - 4	539	90	0.931
H053801		1 - 4	408	86	0.925
H053701		1 - 4	360	84	0.779
H045501		1 - 4	540	78	0.850
H050101		1 - 4	523	86	0.919
H047301		1 - 4	544	88	0.895
H047201		1 - 4	532	85	0.941
H044501		1 - 4	421	93	0.975
H052301		1 - 4	528	92	0.949
H049401		1 - 4	531	88	0.835
H045901		1 - 4	530	81	0.893

Table H-13

*Score Range, Percent Agreement, and Cohen's Kappa**
for the Dichotomously Scored Constructed-Response Geography Items
Used in Main Assessment Scaling, Age 9/Grade 4

Item	Block	Range of Response Codes	Correct Response Codes	Sample Size	Percent Agreement	Cohen's Kappa
G010602		1 - 2	2	570	96	0.93
G010603		1 - 2	2	545	96	0.93
G011401		1 - 2	2	591	98	0.97

Table H-14

*Score Range, Percent Agreement, and Cohen's Kappa**
for the Dichotomously Scored Constructed-Response Geography Items
Used in Main Assessment Scaling, Age 13/Grade 8

Item	Block	Range of Response Codes	Correct Response Codes	Sample Size	Percent Agreement	Cohen's Kappa
G011401		1 - 2	2	552	97	0.94
G015201		1 - 2	2	566	97	0.93

****Note**:* *Cohen's Kappa is a measure of reliability that is appropriate for items that are dichotomized.*

Table H-15

Score Range, Percent Agreement, and Intraclass Correlation for the Polytomously Scored Constructed-Response Geography Items Used in Main Assessment Scaling, Age 9/Grade 4

Item	Block	Range of Response Codes	Sample Size	Percent Agreement	Intraclass Correlation
G012001		1 - 3	408	95	0.96
G013201		1 - 3	424	89	0.86
G011601		1 - 3	543	91	0.95
G007901		1 - 3	572	92	0.93
G008503		1 - 3	560	94	0.95
G007201		1 - 3	571	93	0.96
G010901		1 - 3	523	91	0.94
G007601		1 - 3	541	97	0.98
G008001		1 - 3	579	98	0.99
G012503		1 - 3	586	99	0.99
G012201		1 - 3	601	97	0.98
G009402		1 - 3	608	93	0.93
G009403		1 - 3	588	90	0.91
G010701		1 - 3	574	92	0.94
G010301		1 - 3	577	89	0.91
G009201		1 - 3	609	87	0.92
G008701		1 - 3	572	92	0.95
G011801		1 - 3	460	95	0.96
G012902		1 - 3	569	98	0.99
G010502		1 - 3	566	99	0.99
G007004		1 - 3	565	95	0.98
G011501		1 - 4	565	93	0.96
G009601		1 - 4	563	90	0.97
G007701		1 - 4	743	91	0.96
G013001		1 - 4	547	95	0.98
G008201		1 - 4	574	93	0.96
G009001		1 - 4	540	85	0.94
G007001		1 - 4	786	95	0.98

BEST COPY AVAILABLE

Table H-16

Score Range, Percent Agreement, and Intraclass Correlation for the Polytomously Scored Constructed-response Geography Items Used in Main Assessment Scaling, Age 13/Grade 8

Item	Block	Range of Response Codes	Sample Size	Percent Agreement	Intraclass Correlation
G012001		1 - 3	543	87	0.94
G013201		1 - 3	501	84	0.89
G011601		1 - 3	575	91	0.95
G019402		1 - 3	584	93	0.96
G018801		1 - 3	468	95	0.97
G016502		1 - 3	569	92	0.90
G020001		1 - 3	581	97	0.94
G019003		1 - 3	577	94	0.95
G014401		1 - 3	513	97	0.96
G019202		1 - 3	572	93	0.94
G014001		1 - 3	554	99	1.00
G018101		1 - 3	555	91	0.92
G013402		1 - 3	539	93	0.93
G014301		1 - 3	549	91	0.95
G019102		1 - 3	573	94	0.94
G020302		1 - 3	574	93	0.95
G012503		1 - 3	566	99	0.99
G012201		1 - 3	570	98	0.98
G016201		1 - 3	557	99	0.99
G015901		1 - 3	555	89	0.94
G020201		1 - 3	571	95	0.94
G015301		1 - 3	549	82	0.87
G017701		1 - 3	527	97	0.98
G011801		1 - 3	565	98	0.98
G012902		1 - 3	559	96	0.97
G019002		1 - 3	557	98	0.98
G016302		1 - 3	566	97	0.97
G014901		1 - 3	582	92	0.96
G016401		1 - 3	567	93	0.97
G016701		1 - 3	544	87	0.93
G019302		1 - 3	580	95	0.96
G011501		1 - 4	578	93	0.97
G014201		1 - 4	559	86	0.90
G018301		1 - 4	541	78	0.92
G013001		1 - 4	559	95	0.98
G018501		1 - 4	517	82	0.90
G016001		1 - 4	537	90	0.91
G017101		1 - 4	568	85	0.93
G019901		1 - 4	573	95	0.96

Table H-17

Score Range, Percent Agreement, and Intraclass Correlation for the Polytomously Scored Constructed-Response Geography Items Used in Main Assessment Scaling, Age 17/Grade 12

Item	Block	Range of Response Codes	Sample Size	Percent Agreement	Intraclass Correlation
G020701		1 - 3	528	89	0.94
G018801		1 - 3	458	95	0.97
G026204		1 - 3	536	92	0.96
G026502		1 - 3	530	96	0.97
G016502		1 - 3	532	88	0.91
G025601		1 - 3	511	92	0.89
G018101		1 - 3	522	85	0.92
G026601		1 - 3	539	98	0.98
G022301		1 - 3	541	98	0.99
G023601		1 - 3	523	92	0.95
G016201		1 - 3	516	99	0.99
G024301		1 - 3	484	90	0.95
G025001		1 - 3	522	88	0.90
G017701		1 - 3	538	96	0.97
G022101		1 - 3	515	86	0.88
G026101		1 - 3	559	94	0.96
G024601		1 - 3	272	82	0.84
G025801		1 - 3	369	87	0.88
G016302		1 - 3	538	97	0.98
G025202		1 - 3	497	86	0.91
G021602		1 - 3	496	95	0.96
G016401		1 - 3	524	91	0.96
G016701		1 - 3	527	87	0.92
G026901		1 - 3	532	94	0.98
G021001		1 - 4	512	85	0.89
G026503		1 - 4	533	83	0.93
G025301		1 - 4	531	95	0.95
G021601		1 - 4	504	91	0.94
G018301		1 - 4	525	80	0.92
G021401		1 - 4	532	90	0.98
G018501		1 - 4	527	82	0.91
G023101		1 - 4	493	88	0.89
G022201		1 - 4	516	97	0.98
G017101		1 - 4	539	86	0.95
G023501		1 - 4	531	82	0.93
G025101		1 - 4	511	76	0.90
G026301		1 - 4	538	88	0.83
G022501		1 - 4	529	88	0.95

Appendix I

DIFFERENTIAL ITEM FUNCTIONING (DIF) RESULTS

Appendix I

DIFFERENTIAL ITEM FUNCTIONING (DIF) RESULTS

Table I-1

*1994 Reading Items Identified as "C" or "CC" Items in At Least One Comparison**

Item	Block	Scale	Category	Grade	Comparison	Group Favored
R012511	R10	Reading to Gain Information	C	4	Male/Female	Male
R016208	R13	Reading to Gain Information	C	8	Male/Female	Male
R013406	R10	Reading to Perform a Task	C	8	Male/Female	Female
R013109	R3	Reading for Literary Experience	C	12	Male/Female	Male
R013204	R6	Reading to Gain Information	C	12	Male/Female	Male
R013406	R10	Reading to Perform a Task	C	12	Male/Female	Female
R013914	R11	Reading to Perform a Task	C	12	White/Black	White

*__*Note*__*: For each grade for which an item was administered, three comparisons were performed: Male/Female, White/Black, and White/Hispanic.*

BEST COPY AVAILABLE

Table I-2

*1994 U.S. History Items Identified as "C" or "CC" Items in At Least One Comparison**

Item	Block	Scale	Category	Grade	Comparison	Group Favored
H031001	H5	Democracy	C	4	White/Black	Black
H031301	H5	Democracy	C	4	White/Black	Black
H032401	H6	Technology	C	4	White/Black	White
H033601	H7	Democracy	C	4	White/Black	Black
H036201	H3	Cultures	C	8	White/Black	Black
H038901	H5	Cultures	C	8	White/Black	Black
H040401	H6	Democracy	C	8	White/Black	Black
H037701	H4	World Role	C	8	Male/Female	Male
H051801	H8	Cultures	C	12	White/Black	Black
H040801	H9	Cultures	C	12	White/Black	White
H046801	H4	World Role	C	12	Male/Female	Male
H050901	H7	World Role	C	12	Male/Female	Male
H052801	H8	World Role	C	12	Male/Female	Male
H032503	H6	Democracy	CC	4	White/Black	Black
H032301	H6	Technology	CC	4	Male/Female	Female
H033201	H7	Cultures	CC	8	White/Black	Black
H041801	H9	Democracy	CC	8	White/Black	Black
H038601	H5	Cultures	CC	8	Male/Female	Female
H042902	H10	World Role	CC	8	Male/Female	Male
H040201	H6	Technology	CC	8	White/Hispanic	White
H048202	H5	Technology	CC	12	White/Black	White
H041801	H9	Democracy	CC	12	White/Black	Black
H045501	H4	Cultures	CC	12	Male/Female	Female
H046301	H4	World Role	CC	12	Male/Female	Female
H049503	H6	Cultures	CC	12	Male/Female	Female
H049601	H6	World Role	CC	12	Male/Female	Male
H051101	H7	Cultures	CC	12	Male/Female	Female
H051102	H7	Cultures	CC	12	Male/Female	Female

Note: *For each grade for which an item was administered, three comparisons were performed: Male/Female, White/Black, and White/Hispanic.*

Table I-3

*1994 Geography Items Identified as "C" or "CC" Items in At Least One Comparison**

Item	Block	Scale	Category	Grade	Comparison	Group Favored
G022801	G4	Spatial Dynamics and Connections	C	12	White/Black	Black
G017602	G6	Space and Place	C	12	Male/Female	Male
G016801	G5	Spatial Dynamics and Connections	C	12	White/Hispanic	Hispanic
G010602	G6	Environment and Society	CC	4	White/Hispanic	Hispanic
G018801	G6	Space and Place	CC	8	White/Black	White
G014401	G3	Space and Place	CC	8	Male/Female	Male
G016502	G5	Spatial Dynamics and Connections	CC	8	Male/Female	Female
G012503	G8	Spatial Dynamics and Connections	CC	8	Male/Female	Female
G013201	G8	Environment and Society	CC	8	Male/Female	Female
G019901	G9	Environment and Society	CC	8	White/Hispanic	White
G021401	G3	Space and Place	CC	12	White/Black	White
G016201	G5	Space and Place	CC	12	White/Black	White
G022501	G4	Spatial Dynamics and Connections	CC	12	Male/Female	Female
G010602	G4	Spatial Dynamics and Connections	CC	12	White/Hispanic	Hispanic

****Note**:* *For each grade for which an item was administered, three comparisons were performed: Male/Female, White/Black, and White/Hispanic.*

Appendix J

CORRECTION OF THE NAEP PROGRAM DOCUMENTATION ERROR

Appendix J

CORRECTION OF THE NAEP PROGRAM DOCUMENTATION ERROR

John Mazzeo and Nancy L. Allen
Educational Testing Service

In April 1995, results from the 1994 NAEP reading assessment were released as part of the report *1994 NAEP Reading—A First Look, Findings from the National Assessment of Educational Progress* (Williams, Reese, Campbell, Mazzeo, & Phillips, 1995). Subsequently, ETS/NAEP research scientists discovered an error in the documentation for the ETS version of the PARSCALE program, which is used to compute NAEP scale score results. The error affected how omitted responses were treated in the IRT scaling of the extended constructed-response items that received partial-credit scoring. The error affected only those items; omitted multiple-choice and omitted short constructed responses were treated appropriately.

The conventional treatment in NAEP subjects has been to treat omitted responses (blank responses to an item that are followed by valid responses to items that appear later in the test) as the lowest possible score category in the production of NAEP scale scores. In contrast, not-reached responses (blank responses that are not followed by any further student responses) are treated as missing data. As a result of the documentation error, for a number of the polytomous constructed-response items and across several subject areas, *all* blank responses (both omitted and not-reached responses) to affected items were treated as missing — an *acceptable* treatment but *not* the *conventional* option of choice for NAEP.

The error occurred because of a documentation error in the description of one of the PARSCALE control parameters, designated as POMIT. The program permits the analyst to choose two different ways of treating blank responses: (a) as missing data, and (b) as a valid response falling in the lowest score category. The documentation indicates that by setting POMIT= -1, the treatment in (a) occurs. By setting POMIT=0 or POMIT=1, the treatment in (b) is supposed to occur. The POMIT=1 setting is the program default. In reality, POMIT= 1 and POMIT= -1 operate equivalently, treating blank responses as missing data.

The error appears to have been introduced in 1992 when the programs BILOG and PARSCALE were merged to form the ETS version of PARSCALE. Verification of the accuracy of existing documentation, modifications to internal program diagnostics, and more systematic testing procedures for any and all changes to NAEP-related programs have been implemented immediately to reduce the likelihood of experiencing this kind of error in subsequent NAEP cycles.

The PARSCALE documentation error affected a number of the NAEP scales constructed since 1992, including the 1992 and 1994 national and state reading results. Results from these two assessments have been released to the public in a number of NAEP publications. The 1992 data are also available to the public through NCES's secondary-use data files.

It should be noted that this processing error also impacted the location of the National Assessment Governing Board (NAGB) achievement levels in reading, which were set on the 1992 scales.

NCES and ETS felt that the most technically correct plan of action would be to recalculate all affected NAEP scales, no matter how slight the change, and to issue revised results. ETS was therefore instructed by NCES to recalculate all affected scales and to work with American College Testing (ACT) in the recomputation of the achievement level cutpoints.

In recomputing the cutpoints, an additional error was discovered in the procedures used by ACT in 1992 to "map" the achievement level cutpoints onto the NAEP scale. As described in Appendix K, the procedures contained an incorrectly derived formula. ACT used revised procedures with the correct formula to map the achievement level cutpoints for the 1994 history and geography scales. However, the error in the earlier procedures did affect achievement level cutpoints for reading, which were established during the 1992 assessment. The 1992 national and state reading achievement level results were further impacted by this additional error.

A new version of the 1994 *First Look* report, containing the revised reading results, was issued by NCES in the fall of 1995. The main release of NAEP reading results, including the *1994 NAEP Reading Report Card for the Nation and the States, Findings from the National Assessment of Educational Progress and Trial State Assessments* (Campbell, Donahue, Reese, & Phillips, 1996), *Cross-State Data Compendium for the NAEP 1994 Grade 4 Reading Assessment* (Miller, Nelson, & Naifeh, 1995), individual state reports, almanacs, technical report, and data files, originally scheduled for the end of September 1995, took place instead in January 1996.

The information documenting the original analysis of the 1992 reading data that appears in *The NAEP 1992 Technical Report* (Johnson & Carlson, 1994) is substantially correct for the revised 1992 analysis. Chapter 12 of this technical report describes the revised analysis of the 1994 NAEP reading data, including the revised transformation constants for both 1992 and 1994.

While some *small* changes in scale score results were found, the revised numbers for reading are quite similar to the results released in 1992 and to those published in the NCES April release of the reading *First Look* report. More specifically, the revised reading results are *substantively equivalent* to the originally published 1992 results and to the results released in the *First Look*. Regarding the 1992 and 1994 national assessment data, fourth-grade results are about 1 point lower than originally reported, while twelfth-grade results are about 1 point higher. These changes are small and not substantively meaningful. The eighth-grade numbers are essentially unchanged. The revised numbers indicate the same relative distances between reporting subgroups (i.e., race/ethnicity subgroups, male, females, etc.). The significant national score decline at grade 12 is totally unaffected by the revision, as is the absence of significant changes at grades 4 and 8.

Tables J-1 through J-6 more fully document the effect of the ETS program documentation error on the NAEP reading scale scores for each grade. Tables J-1, J-3, and J-5 contain, respectively for grades 4, 8, and 12, the means, standard deviations, and percentiles for each major reporting subgroup before revisions were made. This information was reported in the first version of the report *1994 NAEP Reading: A First Look* released in April 1995. Tables J-2, J-4, and J-6 contain the same information after revisions were made. This information was reported in the revised *First Look* report and the *Reading Report Card*.

The reanalysis of the 1992 and 1994 reading achievement-level results have been recalculated using two sets of revised achievement-level cutpoints. The first set of revised cutpoints reflects the change in item parameters that resulted from the ETS error. The second set reflects the change in cutpoints due to the ACT error.

In both sets of results, it appears that in virtually all cases there is little difference in the revised and original numbers from an interpretive standpoint. As expected, correction of the ACT error generally results in lower achievement level cutpoints and, hence, slightly higher percentages above the various cutpoints. The revised achievement level results in this technical report and in the reading reports reflect the change in the formula used in setting the achievement levels. Correction of the ETS error impacted results differentially. At grades 4 and 8, small effects were obtained in both directions. At grade 12, correcting the ETS error generally increased the percentages above the cutpoints.

Tables J-7 through J-12 contain information about the effect of the ETS program documentation error and the incorrectly derived "mapping" formula on the achievement level results for each jurisdiction. Tables J-7, J-9, and J-11 contain, respectively for grades 4, 8, and 12, the percentages of students at or above each achievement level and the percentage of student below the Basic level for each major reporting subgroup before revisions were made. These results were reported in the April 1995 version of *1994 NAEP Reading: A First Look, Findings from the National Assessment of Educational Progress*. Tables J-8, J-10, and J-12 contain the percentages for each subgroup after revisions were made. These results were reported in the revised *First Look* report and in the *Reading Report Card*.

Table J-1

NAEP 1992 and 1994 Reading Assessments
Grade 4 Weighted Percentages and Composite Proficiency Means
Weighted Means, Standard Deviations, and Percentiles: Original Results

		MEAN	STD DEV	10TH	25TH	50TH	75TH	90TH
-- TOTAL --	1994	215.2(1.0)	39.9(0.5)	160.6(1.9)<	190.7(1.2)	219.5(1.1)	243.4(1.1)	262.8(1.5)
	1992	217.9(1.0)	36.4(0.6)	169.6(1.6)	194.6(1.2)	220.4(1.1)	243.3(1.1)	262.9(1.5)
SEX								
MALE	1994	210.3(1.2)	41.1(0.7)	153.2(3.4)<	184.7(1.5)	214.8(1.8)	239.3(1.6)	259.3(1.6)
	1992	213.9(1.2)	37.0(0.8)	164.7(2.3)	190.3(1.9)	216.4(1.1)	240.0(1.8)	260.0(1.3)
FEMALE	1994	220.4(1.1)	37.9(0.8)	168.8(2.0)	197.2(1.6)	224.2(1.2)	247.0(1.2)	266.0(1.4)
	1992	222.0(1.0)	35.3(0.7)	175.5(2.3)	200.0(1.2)	224.0(1.1)	246.3(1.7)	265.8(2.8)
REGION								
NORTHEAST	1994	215.6(2.0)	40.7(1.3)	160.6(4.0)	191.6(2.4)	220.2(2.0)	244.4(2.1)	263.7(3.6)
	1992	222.8(3.7)	36.4(1.5)	174.4(4.5)	198.5(3.4)	225.1(3.3)	248.4(2.6)	269.2(5.3)
SOUTHEAST	1994	211.4(2.0)	39.4(1.0)	157.9(4.3)	185.1(2.0)	214.6(2.3)	238.7(2.5)	260.4(1.8)
	1992	213.8(2.4)	36.4(1.2)	165.7(3.3)	190.2(3.2)	215.9(2.0)	238.5(3.4)	258.9(4.3)
CENTRAL	1994	220.6(2.3)	37.4(1.1)	170.4(4.0)	198.6(3.0)	224.8(2.1)	246.3(1.7)	264.3(2.9)
	1992	220.6(1.4)	34.5(1.1)	174.0(3.3)	199.1(1.7)	223.5(1.8)	244.6(1.6)	263.1(2.4)
WEST	1994	213.5(1.9)	41.1(1.1)	155.8(4.3)	188.3(2.6)	218.4(2.6)	242.5(1.9)	262.4(2.3)
	1992	214.9(1.5)	37.6(1.1)	165.3(3.5)	191.3(2.5)	217.3(2.3)	241.3(1.9)	261.1(2.2)
RACE/ETHNICITY								
WHITE	1994	224.8(1.2)	35.3(0.7)	178.5(3.1)	204.1(1.9)	228.0(1.3)	248.7(1.0)	267.3(1.7)
	1992	225.9(1.2)	33.3(0.6)	181.5(2.3)	204.7(1.3)	227.5(1.0)	249.0(1.4)	267.4(1.2)
BLACK	1994	188.5(1.6)	38.4(1.2)	137.9(3.1)	162.7(2.5)	190.2(1.9)	215.5(1.6)	236.5(1.9)
	1992	193.3(1.7)	34.7(1.2)	148.0(3.8)	170.2(2.0)	194.7(2.2)	217.1(1.8)	236.9(2.9)
HISPANIC	1994	192.1(2.5)<	41.2(1.3)	138.4(5.0)	164.0(4.0)	193.8(3.5)	220.9(2.4)	244.2(2.6)
	1992	202.0(2.2)	37.2(1.7)	152.2(5.5)	177.4(2.9)	203.6(3.4)	227.8(2.3)	249.1(2.9)
ASIAN	1994	232.9(5.4)	37.0(3.1)	183.1(6.9)	214.2(8.9)	236.3(2.5)	257.8(7.0)	275.6(4.9)
PACIFIC ISLANDER	1994	220.2(4.9)	38.5(4.8)	161.2(12.0)	199.6(8.6)	226.8(2.4)	245.5(6.3)	264.6(8.9)
AMERICAN INDIAN	1994	202.4(3.4)	39.4(3.9)	148.7(10.9)	176.7(4.4)	206.1(7.3)	230.5(4.2)	250.4(4.2)
	1992	207.5(4.7)	38.6(3.8)	156.5(11.0)	185.0(7.5)	212.0(12.6)	233.2(4.7)	251.4(6.0)

> INDICATES A SIGNIFICANT INCREASE (OR DECREASE "<") BETWEEN 1992 AND 1994

BEST COPY AVAILABLE

Table J-1 (continued)

NAEP 1992 and 1994 Reading Assessments
Grade 4 Weighted Percentages and Composite Proficiency Means
Weighted Means, Standard Deviations, and Percentiles: Original Results

		MEAN	STD DEV	10TH	25TH	50TH	75TH	90TH
PARENTS' EDUCATION LEVEL								
LESS THAN H.S.	1994	189.2(3.3)	37.9(3.1)	139.7(6.3)	164.1(9.7)	189.8(3.2)	216.4(4.0)	236.6(4.6)
	1992	198.8(2.7)	35.0(2.2)	154.3(5.5)	175.6(3.7)	200.3(2.5)	222.7(3.6)	243.3(5.4)
GRADUATED H.S.	1994	208.3(1.8)	39.4(1.8)	152.3(2.2)<	183.7(2.4)	213.4(1.0)	235.8(2.6)	255.4(4.1)
	1992	213.3(1.7)	34.3(1.3)	167.2(2.5)	192.2(1.9)	216.5(1.5)	237.2(1.3)	254.9(2.0)
EDUC AFTER H.S.	1994	224.1(2.0)	36.1(1.2)	175.4(2.9)	202.4(4.1)	228.4(2.1)	248.8(1.9)	266.7(5.1)
	1992	224.0(2.2)	34.4(1.4)	178.9(2.6)	202.8(3.0)	225.4(2.8)	248.0(3.6)	267.0(5.2)
GRAD COLLEGE	1994	225.0(1.2)	38.1(0.8)	172.4(2.0)	202.9(1.8)	229.3(1.4)	251.4(1.2)	270.5(2.3)
	1992	226.7(1.4)	36.1(0.8)	177.7(3.3)	203.4(2.1)	229.5(1.7)	252.4(1.6)	271.1(2.4)
UNKNOWN	1994	207.1(1.2)	39.2(0.7)	153.6(2.6)<	183.0(2.0)	211.2(1.4)	235.3(1.4)	253.8(1.4)
	1992	211.0(1.2)	35.0(0.8)	164.4(2.0)	189.1(1.4)	213.6(1.2)	234.9(1.6)	254.2(2.6)
TYPE OF SCHOOL								
PUBLIC	1994	213.3(1.1)	40.2(0.6)	157.9(2.4)<	188.3(1.6)	217.6(1.2)	241.9(1.2)	261.4(1.4)
	1992	215.9(1.1)	36.5(0.7)	167.6(1.7)	192.6(1.1)	218.3(1.4)	241.3(1.4)	261.2(1.9)
PRIVATE/CATHOLIC	1994	231.6(2.5)	32.8(1.5)	189.3(3.8)	211.6(2.3)	233.6(2.5)	254.2(2.0)	271.7(2.5)
	1992	233.3(1.8)	32.6(0.9)	190.3(3.3)	212.5(1.5)	235.4(1.7)	256.1(1.5)	273.1(2.5)
PRIVATE ONLY	1994	234.6(3.6)	32.2(2.7)	193.3(14.6)	214.6(3.8)	235.2(1.8)	257.1(3.6)	274.2(3.1)
	1992	239.8(3.0)	31.2(1.2)	199.9(6.5)	220.1(6.0)	241.6(3.5)	262.3(3.2)	277.0(1.6)
CATHOLIC ONLY	1994	229.9(3.2)	33.0(1.7)	187.6(3.9)	209.8(3.2)	232.5(2.3)	252.6(2.6)	270.6(3.1)
	1992	230.2(2.2)	32.7(1.2)	186.4(6.7)	209.7(2.3)	232.3(3.7)	252.8(3.2)	270.8(2.6)

> INDICATES A SIGNIFICANT INCREASE (OR DECREASE "<") BETWEEN 1992 AND 1994

BEST COPY AVAILABLE

Table J-2

NAEP 1992 and 1994 Reading Assessments
Grade 4 Weighted Percentages and Composite Proficiency Means
Weighted Means, Standard Deviations, and Percentiles: Revised Results

		MEAN	STD DEV	10TH	25TH	50TH	75TH	90TH
-- TOTAL --	1994	214.3(1.0)	40.6(0.6)	158.6(1.6)<	189.2(1.2)<	218.6(1.1)	242.9(1.1)	262.6(1.5)
	1992	216.7(0.9)	35.6(0.6)	169.5(1.6)	194.0(1.0)	219.2(1.2)	241.6(1.0)	260.9(1.4)
SEX								
MALE	1994	209.2(1.3)	41.8(0.7)	151.1(3.9)<	183.1(1.2)<	213.8(1.8)	238.8(1.4)	259.2(1.7)
	1992	212.8(1.2)	36.2(0.8)	164.8(2.3)	189.7(2.0)	215.2(1.4)	238.3(1.7)	257.9(1.7)
FEMALE	1994	219.5(1.1)	38.5(0.8)	167.0(2.4)	196.0(1.8)	223.3(1.2)	246.6(1.3)	266.0(1.3)
	1992	220.8(1.0)	34.5(0.7)	175.4(2.5)	199.2(1.4)	222.6(1.1)	244.5(1.7)	263.6(2.7)
REGION								
NORTHEAST	1994	214.6(2.1)	41.4(1.3)	158.6(4.3)	190.2(2.5)	219.3(2.0)	244.0(2.4)	263.6(3.6)
	1992	221.5(3.6)	35.6(1.4)	174.4(4.0)	197.8(3.5)	223.8(3.2)	246.6(2.7)	266.9(4.9)
SOUTHEAST	1994	210.4(2.0)	40.1(1.1)	155.8(4.2)	183.6(2.3)	213.6(2.3)	238.2(2.5)	260.3(1.8)
	1992	212.7(2.3)	35.5(1.2)	165.8(3.7)	189.7(2.9)	214.7(1.9)	236.9(3.4)	256.9(3.9)
CENTRAL	1994	219.7(2.4)	38.0(1.1)	168.5(4.3)	197.3(2.8)	223.9(1.9)	246.0(1.9)	264.2(2.7)
	1992	219.4(1.4)	33.7(1.1)	173.8(3.2)	198.4(1.7)	222.2(1.9)	242.7(1.5)	261.1(1.9)
WEST	1994	212.5(2.0)	41.9(1.2)	153.7(4.7)	186.7(2.7)	217.5(2.6)	242.0(1.8)	262.2(2.4)
	1992	213.8(1.4)	36.7(1.0)	165.4(3.2)	190.7(2.5)	216.0(1.9)	239.5(1.9)	258.9(2.0)
RACE/ETHNICITY								
WHITE	1994	224.0(1.3)	36.0(0.7)	176.9(3.6)	203.0(2.3)	227.2(1.6)	248.3(1.1)	267.2(1.5)
	1992	224.6(1.2)	32.6(0.6)	181.2(2.4)	203.8(1.1)	226.1(1.0)	247.1(1.2)	265.2(1.4)
BLACK	1994	187.1(1.7)	39.1(1.2)	135.5(2.8)	160.7(2.1)<	188.7(2.0)	214.6(2.2)	236.0(1.8)
	1992	192.7(1.6)	33.9(1.2)	148.5(3.6)	170.1(2.2)	194.2(2.6)	215.9(2.6)	235.4(3.0)
HISPANIC	1994	190.7(2.6)<	41.9(1.3)	136.2(4.9)	162.1(3.9)<	192.5(3.4)	219.9(2.2)	243.8(2.7)
	1992	201.1(2.1)	36.3(1.6)	152.6(5.7)	177.1(2.8)	202.7(3.2)	226.4(2.0)	247.2(3.3)
ASIAN	1994	232.3(5.5)	37.7(3.1)	181.8(6.7)	213.2(8.9)	235.7(2.5)	257.8(6.5)	275.7(4.7)
PACIFIC ISLANDER	1994	219.4(5.0)	39.2(4.8)	159.3(11.9)	198.4(10.3)	226.1(2.9)	245.2(7.9)	264.7(9.6)
AMERICAN INDIAN	1994	201.2(3.4)	40.1(4.0)	146.5(11.0)	175.1(4.5)	205.0(7.4)	229.9(4.1)	250.2(4.6)
	1992	206.6(4.6)	37.7(3.8)	156.8(11.3)	184.7(9.4)	210.7(8.0)	231.7(4.5)	249.6(5.9)

922

> INDICATES A SIGNIFICANT INCREASE (OR DECREASE "<") BETWEEN 1992 AND 1994

923

Table J-2 (continued)
NAEP 1992 and 1994 Reading Assessments
Grade 4 Weighted Percentages and Composite Proficiency Means
Weighted Means, Standard Deviations, and Percentiles: Revised Results

		MEAN	STD DEV	10TH	25TH	50TH	75TH	90TH
PARENTS' EDUCATION LEVEL								
LESS THAN H.S.	1994	187.8(3.4)	38.6(3.1)	137.4(6.5)	162.0(8.4)	188.5(2.8)	215.5(4.3)	236.0(3.8)
	1992	198.0(2.6)	34.2(2.1)	154.5(5.8)	175.3(4.7)	199.3(2.8)	221.5(3.4)	241.9(2.9)
GRADUATED H.S.	1994	207.2(1.8)	40.1(1.8)	150.3(2.4)<	182.2(2.1)<	212.4(0.8)	235.2(2.3)	255.2(4.0)
	1992	212.2(1.7)	33.5(1.2)	167.2(3.6)	191.7(2.0)	215.4(1.5)	235.7(1.4)	252.8(2.1)
EDUC AFTER H.S.	1994	223.3(2.0)	36.7(1.2)	173.7(4.2)	201.3(5.0)	227.7(2.5)	248.4(2.1)	266.7(4.5)
	1992	222.7(2.1)	33.6(1.3)	178.6(2.1)	202.0(3.1)	224.1(3.2)	246.1(2.8)	264.8(6.5)
GRAD COLLEGE	1994	224.2(1.2)	38.7(0.8)	170.6(2.2)	201.7(1.9)	228.6(1.4)	251.2(1.5)	270.5(2.9)
	1992	225.4(1.4)	35.4(0.8)	177.3(3.2)	202.6(1.7)	228.1(1.6)	250.5(1.6)	268.8(1.7)
UNKNOWN	1994	206.0(1.3)	39.9(0.7)	151.5(3.1)<	181.5(2.0)	210.0(1.3)	234.6(1.4)	253.5(1.4)
	1992	210.0(1.2)	34.2(0.8)	164.4(1.7)	188.7(1.2)	212.5(1.4)	233.3(1.2)	252.1(2.1)
TYPE OF SCHOOL								
PUBLIC	1994	212.3(1.1)	40.9(0.6)	155.9(2.1)<	186.9(1.5)	216.7(1.2)	241.4(1.2)	261.2(1.5)
	1992	214.8(1.0)	35.6(0.6)	167.6(1.9)	192.0(1.0)	217.2(1.7)	239.7(1.3)	259.1(2.3)
PRIVATE/CATHOLIC	1994	231.0(2.5)	33.4(1.5)	187.9(4.2)	210.6(2.5)	232.9(2.3)	253.9(1.9)	271.7(2.7)
	1992	231.8(1.7)	31.8(0.9)	189.8(3.3)	211.5(1.8)	233.8(2.1)	254.1(1.8)	270.8(2.5)
PRIVATE ONLY	1994	234.0(3.7)	32.8(2.8)	192.1(14.8)	213.6(3.6)	234.6(1.7)	256.9(4.0)	274.3(3.5)
	1992	238.2(2.9)	30.5(1.2)	199.1(7.0)	218.8(5.5)	239.9(3.7)	259.8(2.3)	274.7(1.5)
CATHOLIC ONLY	1994	229.2(3.3)	33.5(1.7)	186.1(3.8)	208.6(3.9)	231.9(2.5)	252.2(2.5)	270.5(3.1)
	1992	228.9(2.2)	32.0(1.2)	186.0(6.8)	208.7(2.6)	230.8(3.2)	251.0(3.6)	268.5(2.2)

> INDICATES A SIGNIFICANT INCREASE (OR DECREASE "<") BETWEEN 1992 AND 1994

Table J-3

NAEP 1992 and 1994 Reading Assessments

Grade 8 Weighted Percentages and Composite Proficiency Means

Weighted Means, Standard Deviations, and Percentiles: Original Results

		MEAN	STD DEV	10TH	25TH	50TH	75TH	90TH
-- TOTAL --	1994	260.1(0.8)	36.6(0.4)	211.3(1.9)	236.9(0.9)	262.6(1.1)	285.9(1.2)	305.1(1.3)
	1992	260.2(0.9)	36.0(0.3)	212.6(1.2)	236.7(1.2)	262.3(0.9)	285.3(1.0)	304.8(1.2)
SEX								
MALE	1994	252.8(1.0)	36.8(0.5)	203.7(1.9)	229.4(1.5)	255.3(1.0)	279.0(1.1)	298.0(1.6)
	1992	254.0(1.1)	35.9(0.4)	205.8(1.5)	230.4(1.3)	256.1(1.2)	279.4(1.3)	298.6(1.9)
FEMALE	1994	267.3(0.9)	35.0(0.5)	221.3(2.3)	245.5(1.2)	269.6(1.5)	291.7(1.1)	310.3(1.2)
	1992	266.7(1.0)	34.9(0.5)	221.2(1.7)	243.9(1.2)	268.6(1.2)	291.0(1.2)	309.9(2.2)
REGION								
NORTHEAST	1994	264.9(2.3)	36.5(0.6)	216.1(3.2)	242.0(2.8)	267.9(2.1)	290.3(2.8)	309.6(1.4)
	1992	263.4(1.8)	36.1(1.0)	215.5(1.9)	239.4(2.7)	265.8(1.6)	288.6(1.8)	307.8(2.6)
SOUTHEAST	1994	252.7(1.7)	36.7(0.8)	203.6(3.8)	228.8(2.3)	254.8(1.9)	278.8(1.3)	298.5(1.6)
	1992	254.0(1.7)	36.0(1.0)	206.4(2.7)	230.6(1.9)	255.9(2.0)	279.1(2.6)	299.4(3.2)
CENTRAL	1994	264.8(1.7)	34.5(1.0)	219.2(4.3)	244.0(2.2)	266.9(2.2)	288.8(1.8)	307.0(1.4)
	1992	264.3(2.2)	34.9(0.8)	217.9(4.0)	241.9(2.4)	266.9(2.5)	288.3(1.8)	307.5(2.1)
WEST	1994	259.2(1.2)	37.0(0.7)	210.3(2.7)	236.0(2.0)	261.7(0.9)	285.5(1.7)	304.7(2.0)
	1992	259.6(1.2)	36.1(0.7)	211.4(2.5)	236.5(2.0)	261.7(1.2)	284.7(1.2)	304.2(1.9)
RACE/ETHNICITY								
WHITE	1994	267.9(1.0)	33.6(0.5)	223.8(1.4)	247.4(1.3)	270.0(1.0)	291.1(1.0)	309.0(0.9)
	1992	267.6(1.2)	33.4(0.4)	223.4(2.1)	246.4(1.6)	269.5(1.1)	290.7(1.1)	308.9(1.5)
BLACK	1994	237.1(1.7)	34.4(1.0)	192.2(2.5)	214.4(2.1)	237.8(1.4)	261.0(1.8)	280.6(2.1)
	1992	238.0(1.6)	33.1(0.7)	194.5(3.1)	216.3(1.9)	239.5(1.3)	261.2(2.5)	279.1(1.2)
HISPANIC	1994	240.8(1.4)	37.8(1.0)	190.2(2.7)	216.2(2.0)	242.9(1.7)	267.6(2.1)	288.1(2.5)
	1992	241.6(1.4)	36.5(0.8)	192.6(2.1)	217.4(2.7)	242.5(1.4)	267.0(1.9)	287.5(2.2)
ASIAN	1994	273.8(2.6)	34.3(1.7)	226.2(4.7)	254.0(4.2)	276.3(4.1)	298.1(4.0)	313.9(3.4)
PACIFIC ISLANDER	1994	259.6(7.4)	32.8(2.3)	219.9(6.3)	236.4(14.5)	258.9(6.4)	281.5(3.9)	302.9(4.8)
AMERICAN INDIAN	1994	251.4(4.2)	35.2(2.6)	201.7(13.2)	230.2(3.1)	254.2(4.3)	275.5(7.1)	293.8(5.6)
	1992	251.2(3.7)	34.3(3.4)	208.4(4.9)	230.8(3.2)	252.2(5.0)	274.5(12.0)	292.9(2.8)

> INDICATES A SIGNIFICANT INCREASE (OR DECREASE "<") BETWEEN 1992 AND 1994

BEST COPY AVAILABLE

Table J-3 (continued)

NAEP 1992 and 1994 Reading Assessments
Grade 8 Weighted Percentages and Composite Proficiency Means
Weighted Means, Standard Deviations, and Percentiles: Original Results

		MEAN	STD DEV	10TH	25TH	50TH	75TH	90TH
PARENTS' EDUCATION LEVEL								
LESS THAN H.S.	1994	238.3(1.9)	34.3(1.5)	193.2(4.9)	217.2(3.4)	239.8(1.4)	261.3(3.4)	281.3(2.2)
	1992	242.7(1.5)	34.3(1.7)	197.5(3.0)	221.6(2.2)	243.8(2.2)	266.4(2.3)	286.3(3.6)
GRADUATED H.S.	1994	252.1(1.2)	34.6(1.0)	205.8(4.0)	230.4(1.3)	254.4(1.5)	276.0(1.8)	294.2(1.3)
	1992	250.9(1.4)	34.1(0.7)	205.4(3.1)	228.4(1.8)	253.0(2.0)	275.0(1.3)	293.2(1.8)
EDUC AFTER H.S.	1994	265.9(1.3)	32.7(0.7)	222.8(1.7)	246.0(2.1)	267.9(1.1)	288.3(1.1)	306.3(1.6)
	1992	265.6(1.1)	32.2(0.7)	223.5(1.9)	244.9(1.9)	267.4(1.3)	287.3(1.5)	305.1(2.5)
GRAD COLLEGE	1994	269.9(0.9)	34.8(0.6)	223.1(1.6)	248.5(1.1)	273.0(0.9)	294.3(0.8)	312.0(1.4)
	1992	271.0(1.0)	34.2(0.5)	226.0(1.0)	249.3(1.4)	272.9(1.0)	294.9(1.1)	313.2(0.9)
UNKNOWN	1994	238.7(1.6)	36.7(1.3)	189.4(3.0)	215.3(2.0)	241.5(2.8)	264.3(1.6)	283.7(1.8)
	1992	237.6(2.0)	36.0(1.1)	190.2(3.6)	212.0(3.0)	238.6(2.5)	263.4(2.2)	284.0(1.8)
TYPE OF SCHOOL								
PUBLIC	1994	257.7(0.8)	36.6(0.4)	208.8(1.4)	234.3(1.0)	260.1(0.9)	283.6(0.8)	302.9(1.1)
	1992	258.1(1.0)	35.8(0.3)	210.7(1.6)	234.6(1.2)	260.1(1.1)	283.3(1.0)	302.6(1.1)
PRIVATE/CATHOLIC	1994	279.6(1.4)	29.7(0.8)	241.2(2.3)	261.3(1.6)	280.9(1.5)	299.5(1.4)	316.1(1.8)
	1992	278.5(2.0)	32.2(1.0)	237.1(3.4)	258.4(2.2)	279.9(2.9)	300.4(1.7)	318.4(2.1)
PRIVATE ONLY	1994	280.7(2.4)	29.3(1.2)	242.8(3.4)	262.3(3.7)	281.4(2.5)	299.9(3.7)	316.7(2.6)
	1992	283.4(3.0)	31.7(2.3)	243.5(5.6)	263.5(3.0)	284.7(3.3)	305.2(3.0)	322.0(6.7)
CATHOLIC ONLY	1994	279.0(1.3)	30.0(0.9)	240.1(2.5)	260.6(2.2)	280.5(1.8)	299.1(1.4)	315.4(2.7)
	1992	275.4(1.9)	32.1(0.8)	234.3(2.5)	255.1(1.9)	276.9(1.8)	296.9(1.8)	314.8(2.1)

> INDICATES A SIGNIFICANT INCREASE (OR DECREASE "<") BETWEEN 1992 AND 1994

BEST COPY AVAILABLE

Table J-4

NAEP 1992 and 1994 Reading Assessments
Grade 8 Weighted Percentages and Composite Proficiency Means
Weighted Means, Standard Deviations, and Percentiles: Revised Results

		MEAN	STD DEV	10TH	25TH	50TH	75TH	90TH
-- TOTAL --	1994	259.6(0.8)	36.7(0.4)	210.6(1.4)	236.3(1.1)	262.2(0.7)	285.5(1.1)	304.8(1.1)
	1992	260.0(0.9)	35.9(0.3)	212.5(1.1)	236.6(1.1)	262.1(0.9)	285.1(1.0)	304.5(1.3)
SEX								
MALE	1994	252.4(1.0)	36.9(0.5)	203.1(2.1)	228.9(1.9)	254.9(1.2)	278.6(1.2)	297.7(1.4)
	1992	253.7(1.1)	35.8(0.4)	205.7(1.7)	230.3(1.6)	255.9(1.2)	279.2(1.2)	298.2(1.2)
FEMALE	1994	266.9(1.0)	35.1(0.5)	220.7(2.0)	245.0(1.4)	269.3(1.1)	291.4(1.0)	310.1(1.0)
	1992	266.5(1.0)	34.8(0.5)	221.2(1.5)	243.8(1.2)	268.5(1.2)	290.8(1.0)	309.6(1.7)
REGION								
NORTHEAST	1994	264.5(2.3)	36.7(0.6)	215.7(3.2)	241.5(3.0)	267.5(2.1)	290.0(2.8)	309.4(1.2)
	1992	263.2(1.7)	36.0(1.0)	215.4(2.6)	239.1(3.3)	265.6(1.8)	288.3(1.9)	307.4(2.4)
SOUTHEAST	1994	252.3(1.7)	36.9(0.8)	203.1(3.9)	228.3(2.5)	254.3(1.8)	278.5(1.3)	298.1(1.7)
	1992	253.8(1.7)	35.8(1.0)	206.4(2.4)	230.6(2.1)	255.7(1.9)	278.8(2.6)	299.0(2.9)
CENTRAL	1994	264.4(1.7)	34.7(1.0)	218.8(3.5)	243.6(2.1)	266.6(2.2)	288.4(1.9)	306.7(1.3)
	1992	264.1(2.2)	34.7(0.7)	218.0(3.8)	241.7(2.4)	266.7(2.3)	288.0(1.6)	307.1(2.0)
WEST	1994	258.8(1.2)	37.2(0.7)	209.5(2.2)	235.4(1.5)	261.2(0.9)	285.2(1.8)	304.5(1.9)
	1992	259.4(1.2)	36.1(0.7)	211.3(1.7)	236.3(1.8)	261.5(1.3)	284.5(0.9)	303.8(2.2)
RACE/ETHNICITY								
WHITE	1994	267.5(1.0)	33.7(0.5)	223.3(1.8)	246.9(1.2)	269.7(0.9)	290.8(1.1)	308.7(1.0)
	1992	267.4(1.2)	33.3(0.4)	223.3(1.8)	246.2(1.5)	269.3(1.1)	290.5(1.2)	308.4(1.2)
BLACK	1994	236.5(1.7)	34.6(1.0)	191.5(2.8)	213.7(2.4)	237.3(1.9)	260.6(2.3)	280.2(2.7)
	1992	237.8(1.6)	33.0(0.7)	194.4(2.1)	216.2(2.0)	239.3(1.8)	261.0(2.0)	278.8(1.4)
HISPANIC	1994	240.3(1.4)	38.0(1.0)	189.6(3.1)	215.6(1.7)	242.3(1.3)	267.2(1.8)	287.7(3.3)
	1992	241.4(1.4)	36.4(0.8)	192.6(2.7)	217.2(1.7)	242.4(1.6)	266.8(2.0)	287.1(2.8)
ASIAN	1994	273.4(2.6)	34.4(1.7)	225.5(4.7)	253.5(4.5)	276.1(3.6)	297.8(3.3)	313.6(3.6)
PACIFIC ISLANDER	1994	259.1(7.4)	33.0(2.3)	219.2(6.4)	235.7(14.6)	258.3(6.7)	281.2(3.8)	302.5(5.6)
AMERICAN INDIAN	1994	251.0(4.2)	35.3(2.6)	201.3(13.2)	229.9(3.3)	253.7(4.5)	275.2(7.7)	293.4(7.5)
	1992	251.1(3.7)	34.2(3.4)	208.2(4.6)	231.0(7.1)	252.1(4.9)	274.5(11.2)	292.9(3.3)

> INDICATES A SIGNIFICANT INCREASE (OR DECREASE "<") BETWEEN 1992 AND 1994

Table J-4 (continued)

NAEP 1992 and 1994 Reading Assessments
Grade 8 Weighted Percentages and Composite Proficiency Means
Weighted Means, Standard Deviations, and Percentiles: Revised Results

		MEAN	STD DEV	10TH	25TH	50TH	75TH	90TH
PARENTS' EDUCATION LEVEL								
LESS THAN H.S.	1994	237.8(1.9)	34.5(1.5)	192.6(4.4)	216.6(3.5)	239.4(2.2)	260.9(2.5)	281.1(2.3)
	1992	242.5(1.4)	34.2(1.6)	197.4(3.1)	221.6(2.2)	243.6(2.2)	266.2(2.0)	285.7(3.9)
GRADUATED H.S.	1994	251.6(1.2)	34.7(1.0)	205.2(3.7)	229.8(1.4)	253.9(1.3)	275.5(1.8)	294.0(1.5)
	1992	250.7(1.4)	34.0(0.7)	205.2(4.0)	228.2(1.8)	252.8(1.9)	274.9(1.7)	292.8(1.8)
EDUC AFTER H.S.	1994	265.5(1.3)	32.8(0.7)	222.4(1.4)	245.6(1.7)	267.5(1.1)	287.9(1.0)	306.1(1.6)
	1992	265.3(1.1)	32.1(0.7)	223.1(1.8)	244.7(1.9)	267.2(1.2)	287.1(1.4)	304.9(1.8)
GRAD COLLEGE	1994	269.5(0.9)	35.0(0.6)	222.4(1.6)	248.0(1.2)	272.6(0.9)	294.0(1.0)	311.8(1.4)
	1992	270.8(1.0)	34.0(0.5)	225.8(2.2)	249.0(1.5)	272.7(1.1)	294.6(1.0)	312.8(1.3)
UNKNOWN	1994	238.3(1.6)	36.8(1.3)	188.8(3.3)	214.7(2.5)	241.1(3.3)	263.8(2.1)	283.5(1.9)
	1992	237.5(2.0)	35.9(1.1)	190.3(3.5)	212.1(3.1)	238.5(3.8)	263.1(1.9)	283.9(1.8)
TYPE OF SCHOOL								
PUBLIC	1994	257.3(0.8)	36.8(0.4)	208.3(1.6)	233.8(1.0)	259.7(0.7)	283.2(0.8)	302.6(1.1)
	1992	257.9(1.0)	35.7(0.3)	210.6(2.0)	234.5(1.2)	259.9(1.0)	283.0(1.2)	302.3(1.0)
PRIVATE/CATHOLIC	1994	279.3(1.4)	29.8(0.8)	240.9(2.0)	261.0(1.7)	280.6(1.3)	299.2(1.7)	316.0(2.0)
	1992	278.3(2.0)	32.1(1.0)	237.0(2.6)	258.1(2.3)	279.6(2.9)	300.0(2.4)	318.1(2.9)
PRIVATE ONLY	1994	280.4(2.4)	29.4(1.2)	242.4(3.2)	262.0(3.0)	281.0(2.3)	299.7(3.4)	316.4(2.9)
	1992	283.1(3.0)	31.6(2.3)	243.3(5.6)	263.2(2.7)	284.5(3.3)	304.7(3.4)	321.8(6.5)
CATHOLIC ONLY	1994	278.6(1.3)	30.1(0.9)	239.8(1.8)	260.2(2.7)	280.2(2.3)	298.8(1.8)	315.3(2.4)
	1992	275.2(1.9)	32.0(0.8)	234.2(2.2)	255.0(2.1)	276.8(1.5)	296.6(2.3)	314.4(1.8)

> INDICATES A SIGNIFICANT INCREASE (OR DECREASE "<") BETWEEN 1992 AND 1994

BEST COPY AVAILABLE

Table J-5

NAEP 1992 and 1994 Reading Assessments
Grade 12 Weighted Percentages and Composite Proficiency Means
Weighted Means, Standard Deviations, and Percentiles: Original Results

		MEAN	STD DEV	10TH	25TH	50TH	75TH	90TH
-- TOTAL --	1994	286.5(0.7)<	37.0(0.5)	237.1(1.0)<	263.4(1.1)<	288.9(0.7)<	312.4(0.9)	331.9(1.3)
	1992	291.0(0.6)	33.3(0.4)	246.8(0.7)	269.2(0.8)	292.9(0.8)	314.5(0.6)	332.4(1.0)
SEX								
MALE	1994	279.6(0.8)<	36.6(0.6)	229.9(1.7)<	256.3(1.2)<	282.4(0.9)<	305.7(1.2)	324.5(1.2)
	1992	285.8(0.7)	33.3(0.4)	241.3(1.4)	263.8(1.0)	287.7(1.0)	309.6(1.0)	327.3(0.8)
FEMALE	1994	293.3(0.8)<	36.1(0.7)	246.2(2.0)<	270.6(1.3)	295.5(0.8)	318.5(1.3)	337.7(1.6)
	1992	296.0(0.7)	32.7(0.5)	252.6(1.0)	274.6(1.3)	297.8(1.4)	318.7(1.0)	336.3(1.3)
REGION								
NORTHEAST	1994	287.6(1.7)<	36.8(0.7)	238.8(1.6)	264.5(2.2)	289.1(1.4)<	313.7(2.3)	333.0(1.6)
	1992	293.0(1.2)	34.1(0.8)	246.7(3.3)	270.8(1.9)	295.2(1.2)	317.4(1.2)	334.7(0.9)
SOUTHEAST	1994	281.2(1.2)	36.1(0.8)	233.1(2.0)	258.2(1.3)	282.8(1.2)	306.3(2.6)	326.3(2.0)
	1992	283.5(1.1)	33.3(0.9)	239.5(1.9)	261.5(1.4)	284.7(1.6)	307.0(1.5)	325.5(1.6)
CENTRAL	1994	290.0(1.2)<	36.2(0.9)	242.1(3.8)	268.1(2.0)	293.2(1.6)	315.1(2.1)	333.7(1.9)
	1992	294.4(1.1)	32.1(0.5)	252.1(1.1)	273.7(1.5)	296.1(1.8)	316.8(1.3)	334.0(1.3)
WEST	1994	286.7(1.4)<	38.1(0.8)	235.4(1.9)<	263.0(1.8)<	289.8(1.7)	313.4(1.6)	332.4(2.5)
	1992	292.5(1.6)	33.0(0.8)	249.6(1.5)	270.8(1.8)	294.3(2.1)	315.4(1.9)	333.3(1.3)
RACE/ETHNICITY								
WHITE	1994	293.2(0.7)<	34.6(0.6)	247.7(1.3)<	271.5(0.8)<	295.5(0.8)	317.3(0.9)	335.6(1.5)
	1992	297.0(0.6)	31.4(0.5)	255.6(1.6)	276.9(0.9)	298.7(0.8)	318.8(0.8)	335.6(0.7)
BLACK	1994	263.9(1.6)<	34.3(0.9)	219.0(3.9)	240.9(2.0)<	265.5(2.9)	287.4(2.0)	307.0(2.3)
	1992	271.8(1.5)	31.2(0.8)	231.0(2.1)	250.6(1.5)	272.2(2.0)	293.7(2.0)	312.0(1.4)
HISPANIC	1994	268.5(1.5)<	38.3(1.0)	217.2(3.0)<	243.9(2.5)<	271.1(2.0)	295.2(1.8)	316.2(2.5)
	1992	277.2(2.4)	32.9(1.1)	233.2(2.6)	255.9(2.3)	278.7(2.9)	299.7(3.0)	318.4(3.5)
ASIAN	1994	278.8(2.8)	42.6(2.2)	220.0(9.1)	250.7(4.5)	284.0(4.5)	309.9(2.5)	330.2(1.8)
PACIFIC ISLANDER	1994	279.7(3.9)	36.4(2.7)	230.7(9.0)	259.2(4.6)	282.7(4.3)	304.0(6.9)	323.8(8.7)
AMERICAN INDIAN	1994	273.5(5.4)	33.5(3.9)	231.9(4.0)	249.0(8.2)	273.8(4.1)	295.1(9.0)	317.5(21.3)
	1992	271.8(5.3)	39.4(4.8)	223.2(19.9)	241.2(7.2)	273.9(22.1)	302.5(24.1)	326.8(9.1)

> INDICATES A SIGNIFICANT INCREASE (OR DECREASE "<") BETWEEN 1992 AND 1994

Table J-5 (continued)

NAEP 1992 and 1994 Reading Assessments
Grade 12 Weighted Percentages and Composite Proficiency Means
Weighted Means, Standard Deviations, and Percentiles: Original Results

		MEAN	STD DEV	10TH	25TH	50TH	75TH	90TH
PARENTS' EDUCATION LEVEL								
LESS THAN H.S.	1994	264.7(1.5)<	36.1(1.1)	215.9(3.1)<	240.9(1.3)<	265.9(2.0)	289.5(1.6)	311.2(3.9)
	1992	273.5(1.5)	31.5(0.9)	231.5(2.5)	251.7(2.3)	274.0(2.1)	296.6(2.2)	314.1(1.7)
GRADUATED H.S.	1994	275.8(1.3)<	35.4(0.8)	228.5(4.7)	253.7(1.3)<	277.7(1.7)	300.1(1.9)	319.9(1.3)
	1992	281.4(0.8)	31.7(0.6)	239.9(1.7)	260.1(1.2)	282.2(0.9)	303.7(0.9)	321.6(1.7)
EDUC AFTER H.S.	1994	288.2(1.0)<	33.4(0.8)	244.2(1.7)	267.0(1.7)	290.0(1.1)	311.0(1.3)	330.1(1.9)
	1992	292.7(0.8)	31.1(0.6)	251.4(2.0)	272.3(1.1)	294.4(1.3)	314.4(0.8)	330.8(1.6)
GRAD COLLEGE	1994	297.2(1.0)	34.7(0.7)	251.2(1.8)<	275.7(1.2)<	299.9(0.9)	321.2(1.2)	339.0(2.2)
	1992	300.5(0.8)	31.7(0.5)	258.7(1.4)	280.7(0.9)	302.4(1.0)	322.7(1.1)	339.2(1.0)
UNKNOWN	1994	246.6(2.8)<	36.6(1.8)	198.9(8.9)	223.6(9.5)	247.7(4.3)	271.2(1.4)<	293.7(4.9)
	1992	256.9(2.8)	34.3(2.5)	210.4(10.7)	233.1(5.1)	259.0(7.7)	281.6(2.9)	300.0(5.2)
TYPE OF SCHOOL								
PUBLIC	1994	284.9(0.7)<	37.1(0.6)	235.3(1.1)<	261.7(0.8)<	287.3(0.7)	310.8(0.9)	330.5(1.5)
	1992	288.7(0.7)	33.1(0.4)	244.9(0.9)	266.8(1.0)	290.4(0.9)	312.1(0.8)	330.1(0.9)
PRIVATE/CATHOLIC	1994	300.7(1.9)<	33.1(0.9)	258.1(5.4)	279.6(2.3)<	302.9(1.7)	323.1(2.0)	341.6(1.9)
	1992	306.7(1.3)	30.4(0.6)	268.4(2.1)	288.2(1.2)	308.4(2.3)	327.3(1.5)	344.1(2.0)
PRIVATE ONLY	1994	306.2(2.2)	32.9(1.3)	262.9(2.6)	285.6(3.8)	308.7(3.0)	328.5(3.9)	346.9(2.6)
	1992	307.7(3.0)	34.6(1.6)	261.0(7.9)	286.9(5.6)	310.5(3.3)	331.8(3.6)	349.2(2.7)
CATHOLIC ONLY	1994	297.2(2.4)<	32.7(1.1)	254.2(3.7)<	276.1(2.9)<	299.3(2.7)	319.4(1.5)	337.8(2.1)
	1992	306.3(1.5)	28.1(0.8)	270.6(3.2)	288.6(2.0)	307.3(2.2)	325.3(1.5)	341.6(1.4)

> INDICATES A SIGNIFICANT INCREASE (OR DECREASE "<") BETWEEN 1992 AND 1994

BEST COPY AVAILABLE

Table J-6

NAEP 1992 and 1994 Reading Assessments
Grade 12 Weighted Percentages and Composite Proficiency Means
Weighted Means, Standard Deviations, and Percentiles: Revised Results

		MEAN	STD DEV	10TH	25TH	50TH	75TH	90TH
-- TOTAL --	1994	287.3(0.7)<	36.7(0.5)	238.5(0.9)<	264.4(0.9)<	289.7(0.8)<	313.0(0.9)	332.3(1.3)
	1992	292.1(0.6)	32.8(0.4)	248.7(0.8)	270.6(0.8)	293.9(0.8)	315.3(0.6)	333.0(0.8)
SEX								
MALE	1994	280.4(0.8)<	36.3(0.6)	231.3(1.4)<	257.3(1.2)<	283.1(1.2)<	306.3(1.0)<	324.9(1.3)
	1992	286.9(0.7)	32.7(0.4)	243.3(1.3)	265.1(0.8)	288.7(0.8)	310.3(0.9)	327.7(0.8)
FEMALE	1994	294.2(0.8)<	35.7(0.7)	247.4(1.4)<	271.6(1.4)	296.3(0.8)	319.1(0.9)	338.2(1.7)
	1992	297.1(0.7)	32.1(0.5)	254.6(1.1)	276.0(1.2)	298.8(1.0)	319.5(1.1)	336.9(1.2)
REGION								
NORTHEAST	1994	288.4(1.7)<	36.4(0.7)	240.0(2.0)	265.6(1.9)	289.8(1.7)<	314.3(1.8)	333.5(1.8)
	1992	294.1(1.1)	33.5(0.8)	248.8(2.3)	272.1(2.1)	296.2(1.2)	318.1(1.2)	335.2(1.2)
SOUTHEAST	1994	282.1(1.2)	35.8(0.8)	234.4(1.8)	259.2(1.2)	283.6(1.3)	306.9(2.2)	326.8(1.8)
	1992	284.7(1.1)	32.7(0.9)	241.4(2.1)	262.8(1.3)	285.7(1.7)	307.8(1.3)	326.2(1.7)
CENTRAL	1994	290.8(1.2)<	35.9(0.9)	243.4(3.4)	269.1(2.0)	293.9(1.6)	315.8(2.0)	334.1(2.5)
	1992	295.5(1.1)	31.5(0.5)	253.9(1.3)	275.0(1.5)	297.0(1.8)	317.4(1.2)	334.5(1.4)
WEST	1994	287.6(1.4)<	37.7(0.8)	236.9(2.6)<	264.1(1.7)<	290.5(1.3)	314.0(1.9)	332.7(2.4)
	1992	293.6(1.5)	32.5(0.8)	251.5(1.8)	272.2(1.8)	295.3(2.0)	316.1(1.7)	333.9(1.5)
RACE/ETHNICITY								
WHITE	1994	294.0(0.7)<	34.3(0.6)	248.9(1.1)<	272.5(0.9)<	296.2(0.7)<	317.9(0.9)	336.1(1.4)
	1992	298.0(0.6)	31.0(0.4)	257.2(1.4)	278.1(0.8)	299.6(0.8)	319.5(0.9)	336.1(0.9)
BLACK	1994	264.9(1.6)<	33.9(0.9)	220.4(2.1)<	241.9(2.3)<	266.5(2.0)	288.3(2.0)	307.5(2.5)
	1992	273.2(1.4)	30.6(0.8)	233.3(3.1)	252.6(1.1)	273.4(2.1)	294.7(1.4)	312.8(1.4)
HISPANIC	1994	269.6(1.5)<	37.9(1.0)	218.8(3.1)<	245.5(2.8)<	271.9(1.7)	295.8(2.4)	316.9(2.6)
	1992	278.4(2.3)	32.3(1.1)	235.5(2.5)	257.5(2.0)	279.8(2.9)	300.7(2.5)	319.0(3.8)
ASIAN	1994	279.9(2.8)	42.1(2.2)	222.1(10.9)	251.5(4.8)	284.9(5.1)	310.5(2.7)	331.0(1.8)
PACIFIC ISLANDER	1994	280.4(3.9)	36.0(2.7)	232.3(10.9)	260.0(4.8)	283.3(4.6)	304.5(7.0)	323.9(9.1)
AMERICAN INDIAN	1994	274.6(5.3)	33.2(3.9)	234.1(3.4)	250.3(7.5)	275.2(3.9)	295.9(9.3)	318.3(21.7)
	1992	273.3(5.2)	38.8(4.7)	225.7(18.8)	242.6(8.0)	275.4(26.9)	303.8(30.0)	327.6(9.5)

> INDICATES A SIGNIFICANT INCREASE (OR DECREASE "<") BETWEEN 1992 AND 1994

Table J-6 (continued)

NAEP 1992 and 1994 Reading Assessments
Grade 12 Weighted Percentages and Composite Proficiency Means
Weighted Means, Standard Deviations, and Percentiles: Revised Results

		MEAN	STD DEV	10TH	25TH	50TH	75TH	90TH
PARENTS' EDUCATION LEVEL								
LESS THAN H.S.	1994	265.8(1.5)<	35.7(1.1)	217.3(3.7)	242.2(1.4)<	267.1(1.1)	290.3(1.7)<	311.5(2.1)
	1992	274.9(1.4)	31.0(0.9)	233.9(5.1)	253.5(2.5)	275.4(3.2)	297.6(1.2)	314.8(2.0)
GRADUATED H.S.	1994	276.7(1.3)<	35.0(0.8)	230.0(3.4)	254.9(1.8)	278.6(2.0)	300.7(2.4)	320.3(1.4)
	1992	282.7(0.8)	31.1(0.6)	242.0(2.0)	261.8(1.4)	283.4(1.1)	304.5(0.8)	322.3(1.5)
EDUC AFTER H.S.	1994	289.0(1.0)<	33.2(0.8)	245.3(2.0)	268.0(1.8)	290.8(1.2)	311.7(1.2)	330.6(2.0)
	1992	293.8(0.8)	30.6(0.6)	253.3(1.8)	273.5(1.3)	295.4(1.2)	315.2(0.5)	331.5(1.1)
GRAD COLLEGE	1994	298.0(1.0)<	34.4(0.7)	252.4(1.6)<	276.6(1.2)<	300.5(1.0)	321.8(1.2)	339.5(2.1)
	1992	301.4(0.8)	31.2(0.5)	260.2(1.9)	281.7(0.8)	303.3(1.0)	323.4(1.1)	339.7(1.2)
UNKNOWN	1994	247.9(2.7)<	36.2(1.8)	201.0(7.5)	225.0(5.6)	248.8(3.6)	271.9(1.9)	294.3(3.6)
	1992	258.4(2.8)	33.7(2.4)	213.4(5.0)	234.7(4.6)	261.0(7.9)	282.4(3.0)	300.9(7.9)
TYPE OF SCHOOL								
PUBLIC	1994	285.8(0.7)<	36.8(0.5)	236.7(1.3)<	262.8(0.8)<	288.1(0.7)<	311.5(0.9)	331.0(1.4)
	1992	289.8(0.7)	32.6(0.4)	246.7(0.9)	268.2(0.9)	291.4(0.8)	312.9(0.7)	330.6(0.9)
PRIVATE/CATHOLIC	1994	301.4(1.9)<	32.8(0.9)	259.2(5.2)	280.5(2.4)<	303.4(1.8)	323.6(1.9)	341.9(1.7)
	1992	307.6(1.3)	29.9(0.6)	269.7(2.3)	289.3(1.3)	309.1(1.9)	327.9(1.5)	344.5(1.4)
PRIVATE ONLY	1994	306.9(2.2)	32 5(1.3)	264.0(2.7)	286.3(4.0)	309.3(2.3)	329.1(2.3)	347.2(2.5)
	1992	308.5(2.9)	34.1(1.5)	262.4(8.9)	287.9(5.8)	311.4(3.2)	332.3(3.0)	349.3(1.9)
CATHOLIC ONLY	1994	297.9(2.4)<	32.5(1.1)	255.3(4.3)<	276.8(3.2)<	300.0(3.0)	319.9(1.9)	338.1(1.7)
	1992	307.1(1.5)	27.7(0.7)	272.0(3.2)	289.6(2.0)	308.2(1.7)	325.9(1.2)	341.7(1.5)

> INDICATES A SIGNIFICANT INCREASE (OR DECREASE "<") BETWEEN 1992 AND 1994

BEST COPY AVAILABLE

Table J-7

NAEP 1992 and 1994 Reading Assessments
Grade 4 Weighted Percentages and Composite Proficiency Means
Percent of Students by Achievement Levels: Original Results

		N	WEIGHTED PCT [CV]	MEAN	ADVANCED	PRFCIENT	BASIC	< BASIC
-- TOTAL --	1994	7382	100.0(0.0) [0%]	215.2(1.0)	4.6(0.5)	25.4(1.0)	57.7(1.1)	42.3(1.1)
	1992	6314	100.0(0.0) [0%]	217.9(1.0)	4.5(0.5)	25.3(1.1)	59.0(1.1)	41.0(1.1)
SEX								
MALE	1994	3734	51.2(0.7) [2%]	210.3(1.2)	3.7(0.6)	21.8(1.1)	52.7(1.4)	47.3(1.4)
	1992	3171	51.1(0.6) [1%]	213.9(1.2)	3.6(0.6)	22.4(1.2)	54.5(1.7)	45.5(1.7)
FEMALE	1994	3648	48.8(0.7) [2%]	220.4(1.1)	5.4(0.6)	29.1(1.4)	63.0(1.3)	37.0(1.3)
	1992	3143	48.9(0.6) [1%]	222.0(1.0)	5.5(0.7)	28.4(1.5)	63.7(1.3)	36.3(1.3)
REGION								
NORTHEAST	1994	1816	22.7(0.9) [4%]	215.6(2.0)	4.8(1.2)	26.3(2.2)	58.4(2.0)	41.6(2.0)
	1992	1228	21.4(1.1) [5%]	222.8(3.7)	6.7(2.2)	30.6(4.1)	63.1(3.5)	36.9(3.5)
SOUTHEAST	1994	1888	23.4(1.1) [5%]	211.4(2.0)	4.2(0.6)	21.5(1.7)	52.6(2.3)	47.4(2.3)
	1992	1649	23.4(1.0) [4%]	213.8(2.4)	3.6(0.7)	21.1(2.5)	54.4(3.2)	45.6(3.2)
CENTRAL	1994	1571	24.7(0.7) [3%]	220.6(2.3)	4.9(1.0)	29.0(2.5)	63.6(2.7)	36.4(2.7)
	1992	1599	27.3(0.5) [2%]	220.6(1.4)	4.2(0.9)	26.5(2.1)	63.4(2.0)	36.6(2.0)
WEST	1994	2107	29.2(0.8) [3%]	213.5(1.9)	4.4(0.6)	24.7(1.7)	56.4(2.3)	43.6(2.3)
	1992	1838	28.0(0.8) [3%]	214.9(1.5)	4.0(0.6)	23.6(1.4)	55.6(1.8)	44.4(1.8)
RACE/ETHNICITY								
WHITE	1994	4401	68.7(0.2) [0%]	224.8(1.2)	5.9(0.7)	31.5(1.3)	68.1(1.3)	31.9(1.3)
	1992	3917	70.6(0.2) [1%]	225.9(1.2)	6.0(0.7)	31.4(1.6)	68.1(1.4)	31.9(1.4)
BLACK	1994	1261	15.0(0.2) [1%]	188.5(1.6)	0.6(0.3)	7.2(0.9)	28.4(2.2)	71.6(2.2)
	1992	1013	15.8(0.1) [1%]	193.3(1.7)	0.4(0.2)	7.0(1.4)	30.9(2.3)	69.1(2.3)
HISPANIC	1994	1184	11.6(0.2) [2%]	192.1(2.5)<	1.3(0.5)	10.8(1.6)	32.9(2.6)	67.1(2.6)
	1992	1044	9.4(0.1) [2%]	202.0(2.2)	1.7(0.6)	13.4(1.8)	41.0(2.2)	59.0(2.2)
ASIAN	1994	211	1.9(0.2) [10%]	232.9(5.4)	9.9(2.5)	42.8(7.3)	76.7(5.7)	23.3(5.7)
PACIFIC ISLANDER	1994	96	0.7(0.1) [17%]	220.2(4.9)	4.8(3.5)	29.3(4.8)	63.6(6.4)	36.4(6.4)
AMERICAN INDIAN	1994	170	1.5(0.2) [10%]	202.4(3.4)	1.9(1.6)	14.6(3.3)	44.8(4.6)	55.2(4.6)
	1992	117	1.7(0.2) [15%]	207.5(4.7)	2.4(1.9)	15.4(4.7)	50.5(6.1)	49.5(6.1)

> INDICATES A SIGNIFICANT INCREASE (OR DECREASE "<") BETWEEN 1992 AND 1994

942

943

Table J-7 (continued)

NAEP 1992 and 1994 Reading Assessments
Grade 4 Weighted Percentages and Composite Proficiency Means
Percent of Students by Achievement Levels: Original Results

		N	WEIGHTED PCT [CV]	MEAN	ADVANCED	PRFCIENT	BASIC	< BASIC
PARENTS' EDUCATION LEVEL								
LESS THAN H.S.	1994	303	4.1(0.3) [8%]	189.2(3.3)	0.8(1.1)	7.1(2.6)	29.1(3.6)	70.9(3.6)
	1992	291	4.1(0.4) [9%]	198.8(2.7)	1.1(1.3)	10.4(2.4)	34.7(3.8)	65.3(3.8)
GRADUATED H.S.	1994	889	12.5(0.5) [4%]	208.3(1.8)	2.1(0.9)	18.9(2.5)	51.5(1.9)	48.5(1.9)
	1992	769	12.4(0.6) [5%]	213.3(1.7)	2.3(1.0)	18.7(2.2)	54.8(2.5)	45.2(2.5)
EDUC AFTER H.S.	1994	540	7.6(0.5) [6%]	224.1(2.0)	5.2(1.2)	32.1(2.7)	66.6(2.7)	33.4(2.7)
	1992	529	8.6(0.5) [6%]	224.0(2.2)	5.9(1.9)	29.4(3.0)	66.4(3.0)	33.6(3.0)
GRAD COLLEGE	1994	3134	41.9(0.9) [2%]	225.0(1.2)	7.5(1.1)	34.0(1.6)	67.9(1.3)	32.1(1.3)
	1992	2481	39.0(1.1) [3%]	226.7(1.4)	7.6(0.8)	35.0(1.7)	68.1(1.8)	31.9(1.8)
UNKNOWN	1994	2501	33.6(0.8) [2%]	207.1(1.2)	2.1(0.3)	17.9(1.5)	49.1(1.4)	50.9(1.4)
	1992	2228	35.7(1.0) [3%]	211.0(1.2)	2.1(0.5)	18.0(1.2)	52.0(1.7)	48.0(1.7)
TYPE OF SCHOOL								
PUBLIC	1994	6030	89.7(0.9) [1%]	213.3(1.1)	4.1(0.5)	24.0(1.1)	55.8(1.2)	44.2(1.2)
	1992	5045	88.1(1.3) [1%]	215.9(1.1)	4.0(0.6)	23.5(1.2)	56.9(1.2)	43.1(1.2)
PRIVATE/CATHOLIC	1994	1352	10.3(0.9) [9%]	231.6(2.5)	8.1(1.3)	37.6(2.7)	74.6(2.5)	25.4(2.5)
	1992	1241	11.1(1.0) [9%]	233.3(1.8)	8.8(1.3)	40.2(2.6)	75.7(1.9)	24.3(1.9)
PRIVATE ONLY	1994	442	3.8(0.6) [16%]	234.6(3.6)	9.6(2.3)	41.2(3.7)	77.5(4.2)	22.5(4.2)
	1992	327	3.5(0.7) [21%]!	239.8(3.0)	11.6(2.3)	48.6(4.9)	82.3(2.7)	17.7(2.7)
CATHOLIC ONLY	1994	910	6.5(0.8) [13%]	229.9(3.2)	7.2(1.8)	35.5(3.7)	72.8(3.6)	27.2(3.6)
	1992	914	7.6(0.8) [10%]	230.2(2.2)	7.4(1.5)	36.2(2.7)	72.6(2.5)	27.4(2.5)

! INTERPRET WITH CAUTION: SAMPLING ERROR CANNOT BE ESTIMATED ACCURATELY SINCE COEFFICIENT OF VARIATION OF ESTIMATED NUMBER OF STUDENTS EXCEEDS 20%

> INDICATES A SIGNIFICANT INCREASE (OR DECREASE "<") BETWEEN 1992 AND 1994

Table J-8

NAEP 1992 and 1994 Reading Assessments
Grade 4 Weighted Percentages and Composite Proficiency Means
Percent of Students by Achievement Levels: Revised Results

		N	WEIGHTED PCT [CV]	MEAN	ADVANCED	PRFCIENT	BASIC	< BASIC
-- TOTAL --	1994	7382	100.0(0.0) [0%]	214.3(1.0)	7.4(0.7)	29.6(1.1)	60.5(1.0)	39.5(1.0)
	1992	6314	100.0(0.0) [0%]	216.7(0.9)	6.4(0.6)	28.6(1.2)	62.1(1.1)	37.9(1.1)
SEX								
MALE	1994	3734	51.2(0.7) [2%]	209.2(1.3)	6.1(0.8)	25.8(1.3)	55.4(1.4)	44.6(1.4)
	1992	3171	51.1(0.6) [1%]	212.8(1.2)	5.2(0.7)	25.4(1.4)	57.6(1.6)	42.4(1.6)
FEMALE	1994	3648	48.8(0.7) [2%]	219.5(1.1)	8.7(0.9)	33.6(1.5)	65.8(1.2)	34.2(1.2)
	1992	3143	48.9(0.6) [1%]	220.8(1.0)	7.7(0.8)	31.9(1.4)	66.7(1.3)	33.3(1.3)
REGION								
NORTHEAST	1994	1816	22.7(0.9) [4%]	214.6(2.1)	7.9(1.4)	30.6(2.4)	60.9(2.1)	39.1(2.1)
	1992	1228	21.4(1.1) [5%]	221.5(3.6)	9.4(2.4)	33.8(4.3)	66.0(3.6)	34.0(3.6)
SOUTHEAST	1994	1888	23.4(1.1) [5%]	210.4(2.0)	6.7(0.9)	25.3(2.1)	55.2(2.3)	44.8(2.3)
	1992	1649	23.4(1.0) [4%]	212.7(2.3)	4.9(1.0)	23.9(2.6)	57.5(3.1)	42.5(3.1)
CENTRAL	1994	1571	24.7(0.7) [3%]	219.7(2.4)	7.7(1.1)	33.7(2.5)	66.4(2.6)	33.6(2.6)
	1992	1599	27.3(0.5) [2%]	219.4(1.4)	6.1(1.1)	30.4(2.1)	66.4(1.7)	33.6(1.7)
WEST	1994	2107	29.2(0.8) [3%]	212.5(2.0)	7.3(0.8)	29.0(1.8)	59.3(2.1)	40.7(2.1)
	1992	1838	28.0(0.8) [3%]	213.8(1.4)	5.6(0.7)	26.6(1.7)	58.6(1.7)	41.4(1.7)
RACE/ETHNICITY								
WHITE	1994	4401	68.7(0.2) [0%]	224.0(1.3)	9.5(0.9)	36.6(1.4)	70.8(1.2)	29.2(1.2)
	1992	3917	70.6(0.2) [1%]	224.6(1.2)	8.4(0.9)	35.2(1.7)	71.2(1.3)	28.8(1.3)
BLACK	1994	1261	15.0(0.2) [1%]	187.1(1.7)	1.1(0.4)	9.0(1.0)	31.3(2.5)	68.7(2.5)
	1992	1013	15.8(0.1) [1%]	192.7(1.6)	0.7(0.4)	8.4(1.4)	33.3(2.3)	66.7(2.3)
HISPANIC	1994	1184	11.6(0.2) [2%]	190.7(2.6)<	2.4(0.6)	13.1(1.6)	35.6(2.6)	64.4(2.6)
	1992	1044	9.4(0.1) [2%]	201.1(2.1)	2.5(0.8)	15.5(1.8)	44.0(2.2)	56.0(2.2)
ASIAN	1994	211	1.9(0.2) [10%]	232.3(5.5)	15.6(5.7)	47.9(7.1)	78.3(5.1)	21.7(5.1)
PACIFIC ISLANDER	1994	96	0.7(0.1) [17%]	219.4(5.0)	7.6(3.6)	34.5(4.6)	66.6(6.9)	33.4(6.9)
AMERICAN INDIAN	1994	170	1.5(0.2) [10%]	201.2(3.4)	3.1(2.1)	18.1(3.8)	47.7(4.4)	52.3(4.4)
	1992	117	1.7(0.2) [15%]	206.6(4.6)	3.4(2.1)	18.2(4.5)	53.2(6.6)	46.8(6.6)

> INDICATES A SIGNIFICANT INCREASE (OR DECREASE "<") BETWEEN 1992 AND 1994

Table J-8 (continued)

NAEP 1992 and 1994 Reading Assessments

Grade 4 Weighted Percentages and Composite Proficiency Means

Percent of Students by Achievement Levels: Revised Results

		N	WEIGHTED PCT [CV]	MEAN	ADVANCED	PRFCIENT	BASIC	< BASIC
PARENTS' EDUCATION LEVEL								
LESS THAN H.S.	1994	303	4.1(0.3) [8%]	187.8(3.4)	1.3(1.3)	8.8(2.3)	31.6(3.9)	68.4(3.9)
	1992	291	4.1(0.4) [9%]	198.0(2.6)	1.4(1.4)	11.8(2.2)	38.8(3.8)	61.2(3.8)
GRADUATED H.S.	1994	889	12.5(0.5) [4%]	207.2(1.8)	4.0(1.3)	22.4(2.6)	54.5(2.1)	45.5(2.1)
	1992	769	12.4(0.6) [5%]	212.2(1.7)	3.1(1.1)	22.1(2.2)	57.7(2.2)	42.3(2.2)
EDUC AFTER H.S.	1994	540	7.6(0.5) [6%]	223.3(2.0)	9.4(1.8)	37.1(2.6)	69.5(2.9)	30.5(2.9)
	1992	529	8.6(0.5) [6%]	222.7(2.1)	8.1(1.9)	33.1(3.4)	69.5(3.0)	30.5(3.0)
GRAD COLLEGE	1994	3134	41.9(0.9) [2%]	224.2(1.2)	11.4(1.2)	39.0(1.7)	70.4(1.3)	29.6(1.3)
	1992	2481	39.0(1.1) [3%]	225.4(1.4)	10.5(1.1)	38.7(1.8)	70.8(1.6)	29.2(1.6)
UNKNOWN	1994	2501	33.6(0.8) [2%]	206.0(1.3)	4.0(0.6)	21.7(1.6)	52.1(1.3)	47.9(1.3)
	1992	2228	35.7(1.0) [3%]	210.0(1.2)	3.3(0.8)	20.7(1.4)	55.4(1.6)	44.6(1.6)
TYPE OF SCHOOL								
PUBLIC	1994	6030	89.7(0.9) [1%]	212.3(1.1)	6.8(0.7)	28.1(1.2)	58.6(1.1)	41.4(1.1)
	1992	5045	88.1(1.3) [1%]	214.8(1.0)	5.7(0.6)	26.6(1.3)	60.0(1.1)	40.0(1.1)
PRIVATE/CATHOLIC	1994	1352	10.3(0.9) [9%]	231.0(2.5)	12.7(1.8)	43.2(3.0)	77.1(2.4)	22.9(2.4)
	1992	1241	11.1(1.0) [9%]	231.8(1.7)	11.9(1.3)	44.5(2.4)	78.6(1.9)	21.4(1.9)
PRIVATE ONLY	1994	442	3.8(0.6) [16%]	234.0(3.7)	14.3(2.9)	46.0(4.0)	79.7(4.2)	20.3(4.2)
	1992	327	3.5(0.7) [21%]!	238.2(2.9)	15.3(2.9)	52.9(4.4)	84.4(2.7)	15.6(2.7)
CATHOLIC ONLY	1994	910	6.5(0.8) [13%]	229.2(3.3)	11.8(2.2)	41.6(3.9)	75.6(3.2)	24.4(3.2)
	1992	914	7.6(0.8) [10%]	228.9(2.2)	10.3(1.5)	40.6(2.7)	75.8(2.7)	24.2(2.7)

! INTERPRET WITH CAUTION: SAMPLING ERROR CANNOT BE ESTIMATED ACCURATELY SINCE COEFFICIENT OF VARIATION OF ESTIMATED NUMBER OF STUDENTS EXCEEDS 20%

> INDICATES A SIGNIFICANT INCREASE (OR DECREASE "<") BETWEEN 1992 AND 1994

BEST COPY AVAILABLE

Table J-9

NAEP 1992 and 1994 Reading Assessments, Grade 8 Weighted Percentages and Composite Proficiency Means, Percent of Students by Achievement Levels: Original Results

		N	WEIGHTED PCT [CV]	MEAN	ADVANCED	PRFCIENT	BASIC	< BASIC
-- TOTAL --	1994	10135	100.0(0.0) [0%]	260.1(0.8)	1.9(0.2)	27.8(0.9)	69.1(0.9)	30.9(0.9)
	1992	9464	100.0(0.0) [0%]	260.2(0.9)	2.1(0.3)	27.5(1.1)	68.8(1.0)	31.2(1.0)
SEX								
MALE	1994	5009	50.0(0.6) [1%]	252.8(1.0)	1.0(0.2)	21.1(1.0)	62.0(1.1)	38.0(1.1)
	1992	4689	50.8(0.7) [1%]	254.0(1.1)	1.1(0.3)	21.8(1.2)	62.8(1.2)	37.2(1.2)
FEMALE	1994	5126	50.0(0.6) [1%]	267.3(0.9)	2.8(0.4)	34.6(1.2)	76.2(1.1)	23.8(1.1)
	1992	4775	49.2(0.7) [1%]	266.7(1.0)	3.1(0.5)	33.4(1.4)	74.9(1.1)	25.1(1.1)
REGION								
NORTHEAST	1994	1918	20.4(0.8) [4%]	264.9(2.3)	2.9(0.7)	32.8(2.7)	73.5(2.1)	26.5(2.1)
	1992	2016	22.0(0.7) [3%]	263.4(1.8)	2.7(0.4)	31.1(1.9)	71.0(2.3)	29.0(2.3)
SOUTHEAST	1994	3132	25.9(1.0) [4%]	252.7(1.7)	0.9(0.3)	21.1(1.4)	61.2(1.9)	38.8(1.9)
	1992	2456	24.8(0.5) [2%]	254.0(1.7)	1.1(0.4)	21.6(2.3)	62.8(1.8)	37.2(1.8)
CENTRAL	1994	2149	24.2(0.6) [3%]	264.8(1.7)	2.2(0.8)	31.3(2.0)	75.0(1.9)	25.0(1.9)
	1992	2460	25.3(0.5) [2%]	264.3(2.2)	2.5(0.6)	30.7(2.4)	73.1(2.4)	26.9(2.4)
WEST	1994	2936	29.5(0.8) [3%]	259.2(1.2)	1.8(0.4)	27.5(1.2)	68.3(1.3)	31.7(1.3)
	1992	2532	27.9(0.6) [2%]	259.6(1.2)	2.1(0.5)	27.1(1.4)	68.4(1.5)	31.6(1.5)
RACE/ETHNICITY								
WHITE	1994	6312	69.5(0.3) [1%]	267.9(1.0)	2.4(0.4)	34.1(1.2)	77.8(1.1)	22.2(1.1)
	1992	5994	69.9(0.2) [0%]	267.6(1.2)	2.6(0.4)	33.7(1.5)	76.8(1.1)	23.2(1.1)
BLACK	1994	1685	15.1(0.2) [1%]	237.1(1.7)	0.2(0.2)	8.5(1.2)	43.3(1.9)	56.7(1.9)
	1992	1568	15.5(0.2) [1%]	238.0(1.6)	0.1(0.2)	8.0(1.0)	44.4(1.9)	55.6(1.9)
HISPANIC	1994	1457	10.6(0.2) [1%]	240.8(1.4)	0.4(0.2)	13.3(1.4)	48.9(1.7)	51.1(1.7)
	1992	1402	10.2(0.2) [2%]	241.6(1.4)	0.6(0.3)	12.7(1.1)	48.7(2.1)	51.3(2.1)
ASIAN	1994	290	1.9(0.2) [8%]	273.8(2.6)	4.3(1.2)	42.3(3.5)	80.4(2.9)	19.6(2.9)
PACIFIC ISLANDER	1994	178	1.3(0.4) [28%]!	259.6(7.4)	2.0(2.7)	24.8(7.8)	67.1(10.3)	32.9(10.3)
AMERICAN INDIAN	1994	159	1.3(0.2) [19%]	251.4(4.2)	0.3(0.4)	18.7(5.4)	62.0(5.7)	38.0(5.7)
	1992	106	1.1(0.2) [14%]	251.2(3.7)	0.6(0.9)	18.4(7.2)	60.3(5.0)	39.7(5.0)

! INTERPRET WITH CAUTION: SAMPLING ERROR CANNOT BE ESTIMATED ACCURATELY SINCE COEFFICIENT OF VARIATION OF ESTIMATED NUMBER OF STUDENTS EXCEEDS 20%

> INDICATES A SIGNIFICANT INCREASE (OR DECREASE "<") BETWEEN 1992 AND 1994

Table J-9 (continued)

NAEP 1992 and 1994 Reading Assessments, Grade 8 Weighted Percentages and Composite Proficiency Means, Percent of Students by Achievement Levels: Original Results

		N	WEIGHTED PCT [CV	MEAN	ADVANCED	PRFCIENT	BASIC	< BASIC
PARENTS' EDUCATION LEVEL								
LESS THAN H.S.	1994	732	6.9(0.4) [6%]	238.3(1.9)	0.2(0.2)	9.1(1.8)	45.7(2.8)	54.3(2.8)
	1992	746	7.9(0.5) [6%]	242.7(1.5)	0.4(0.1)	11.7(1.8)	49.8(2.3)	50.2(2.3)
GRADUATED H.S.	1994	2086	21.4(0.8) [4%]	252.1(1.2)	0.7(0.4)	18.5(1.4)	61.4(1.6)	38.6(1.6)
	1992	2137	23.7(0.8) [3%]	250.9(1.4)	0.7(0.2)	17.4(1.6)	59.7(1.8)	40.3(1.8)
EDUC AFTER H.S.	1994	1926	19.5(0.5) [3%]	265.9(1.3)	1.9(0.6)	31.0(1.5)	76.8(1.5)	23.2(1.5)
	1992	1806	19.1(0.5) [2%]	265.6(1.1)	2.2(0.7)	30.6(1.6)	75.6(1.5)	24.4(1.5)
GRAD COLLEGE	1994	4378	42.8(1.1) [2%]	269.9(0.9)	3.1(0.4)	38.0(1.1)	78.4(1.0)	21.6(1.0)
	1992	3946	41.3(1.2) [3%]	271.0(1.0)	3.5(0.5)	38.1(1.4)	79.3(1.0)	20.7(1.0)
UNKNOWN	1994	966	8.9(0.4) [4%]	238.7(1.6)	0.3(0.3)	10.5(1.4)	47.2(2.7)	52.8(2.7)
	1992	805	7.8(0.4) [6%]	237.6(2.0)	0.3(0.3)	11.1(1.9)	44.3(2.5)	55.7(2.5)
TYPE OF SCHOOL								
PUBLIC	1994	8327	89.4(1.0) [1%]	257.7(0.8)	1.6(0.3)	25.6(0.9)	66.9(0.9)	33.1(0.9)
	1992	7656	89.4(0.8) [1%]	258.1(1.0)	1.7(0.3)	25.3(1.1)	66.7(1.1)	33.3(1.1)
PRIVATE/CATHOLIC	1994	1808	10.6(1.0) [9%]	279.6(1.4)	4.3(0.7)	47.0(2.3)	88.3(1.3)	11.7(1.3)
	1992	1808	10.6(0.8) [8%]	278.5(2.0)	5.3(1.3)	46.2(3.1)	86.1(1.6)	13.9(1.6)
PRIVATE ONLY	1994	672	4.0(0.7) [17%]	280.7(2.4)	4.8(1.1)	47.6(4.0)	89.1(2.1)	10.9(2.1)
	1992	621	4.1(0.8) [19%]	283.4(3.0)	6.8(2.2)	51.9(4.8)	89.7(2.6)	10.3(2.6)
CATHOLIC ONLY	1994	1136	6.6(0.6) [10%]	279.0(1.3)	4.1(1.0)	46.7(2.1)	87.8(1.3)	12.2(1.3)
	1992	1187	6.5(0.6) [10%]	275.4(1.9)	4.4(1.0)	42.6(2.7)	83.8(1.6)	16.2(1.6)

! INTERPRET WITH CAUTION: SAMPLING ERROR CANNOT BE ESTIMATED ACCURATELY SINCE COEFFICIENT OF VARIATION OF ESTIMATED NUMBER OF STUDENTS EXCEEDS 20%

> INDICATES A SIGNIFICANT INCREASE (OR DECREASE "<") BETWEEN 1992 AND 1994

BEST COPY AVAILABLE

Table J-10

NAEP 1992 and 1994 Reading Assessments, Grade 8 Weighted Percentages and Composite Proficiency Means, Percent of Students by Achievement Levels: Revised Results

		N	WEIGHTED PCT [CV]	MEAN	ADVANCED	PRFCIENT	BASIC	< BASIC
-- TOTAL --	1994	10135	100.0(0.0) [0%]	259.6(0.8)	2.8(0.3)	29.5(0.9)	69.6(0.9)	30.4(0.9)
	1992	9464	100.0(0.0) [0%]	260.0(0.9)	2.9(0.3)	29.2(1.1)	69.5(1.0)	30.5(1.0)
SEX								
MALE	1994	5009	50.0(0.6) [1%]	252.4(1.0)	1.5(0.3)	22.7(1.1)	62.5(1.1)	37.5(1.1)
	1992	4689	50.8(0.7) [1%]	253.7(1.1)	1.7(0.2)	23.2(1.2)	63.5(1.3)	36.5(1.3)
FEMALE	1994	5126	50.0(0.6) [1%]	266.9(1.0)	4.1(0.5)	36.4(1.3)	76.6(1.1)	23.4(1.1)
	1992	4775	49.2(0.7) [1%]	266.5(1.0)	4.2(0.6)	35.3(1.4)	75.6(1.1)	24.4(1.1)
REGION								
NORTHEAST	1994	1918	20.4(0.8) [4%]	264.5(2.3)	4.0(0.8)	34.6(2.7)	74.0(2.2)	26.0(2.2)
	1992	2016	22.0(0.7) [3%]	263.2(1.7)	3.7(0.6)	33.0(2.1)	71.7(2.2)	28.3(2.2)
SOUTHEAST	1994	3132	25.9(1.0) [4%]	252.3(1.7)	1.6(0.4)	22.6(1.4)	61.7(1.9)	38.3(1.9)
	1992	2456	24.8(0.5) [2%]	253.8(1.7)	1.7(0.5)	23.0(2.5)	63.6(1.8)	36.4(1.8)
CENTRAL	1994	2149	24.2(0.6) [3%]	264.4(1.7)	3.3(0.7)	33.2(2.2)	75.3(1.7)	24.7(1.7)
	1992	2460	25.3(0.5) [2%]	264.1(2.2)	3.5(0.7)	32.5(2.4)	73.9(2.3)	26.1(2.3)
WEST	1994	2936	29.5(0.8) [3%]	258.8(1.2)	2.7(0.4)	29.1(1.3)	68.7(1.3)	31.3(1.3)
	1992	2532	27.9(0.6) [2%]	259.4(1.2)	2.9(0.6)	28.6(1.5)	69.0(1.5)	31.0(1.5)
RACE/ETHNICITY								
WHITE	1994	6312	69.5(0.3) [1%]	267.5(1.0)	3.6(0.4)	36.1(1.3)	78.2(1.1)	21.8(1.1)
	1992	5994	69.9(0.2) [0%]	267.4(1.2)	3.7(0.5)	35.6(1.5)	77.5(1.1)	22.5(1.1)
BLACK	1994	1685	15.1(0.2) [1%]	236.5(1.7)	0.4(0.3)	9.4(1.2)	43.7(1.9)	56.3(1.9)
	1992	1568	15.5(0.2) [1%]	237.8(1.6)	0.2(0.2)	8.9(1.1)	45.5(1.8)	54.5(1.8)
HISPANIC	1994	1457	10.6(0.2) [1%]	240.3(1.4)	0.6(0.3)	14.2(1.5)	49.4(1.6)	50.6(1.6)
	1992	1402	10.2(0.2) [2%]	241.4(1.4)	0.8(0.3)	14.0(1.3)	49.3(2.2)	50.7(2.2)
ASIAN	1994	290	1.9(0.2) [8%]	273.4(2.6)	5.9(1.8)	44.4(3.7)	80.6(2.9)	19.4(2.9)
PACIFIC ISLANDER	1994	178	1.3(0.4) [20%]!	259.1(7.4)	3.0(3.1)	25.6(8.1)	67.5(9.9)	32.5(9.9)
AMERICAN INDIAN	1994	159	1.3(0.2) [19%]	251.0(4.2)	1.0(1.1)	20.2(5.6)	62.8(5.6)	37.2(5.6)
	1992	106	1.1(0.2) [14%]	251.1(3.7)	0.6(0.9)	19.8(7.3)	60.7(5.0)	39.3(5.0)

! INTERPRET WITH CAUTION: SAMPLING ERROR CANNOT BE ESTIMATED ACCURATELY SINCE COEFFICIENT OF VARIATION OF ESTIMATED NUMBER OF STUDENTS EXCEEDS 20%

> INDICATES A SIGNIFICANT INCREASE (OR DECREASE "<") BETWEEN 1992 AND 1994

Table J-10 (continued)

NAEP 1992 and 1994 Reading Assessments, Grade 8 Weighted Percentages and Composite Proficiency Means, Percent of Students by Achievement Levels: Revised Results

		N	WEIGHTED PCT [CV]	MEAN	ADVANCED	PRFCIENT	BASIC	< BASIC
PARENTS' EDUCATION LEVEL								
LESS THAN H.S.	1994	732	6.9(0.4) [6%]	237.8(1.9)	0.3(0.3)	10.0(1.8)	46.1(3.0)	53.9(3.0)
	1992	746	7.9(0.5) [6%]	242.5(1.4)	0.6(0.3)	13.1(1.9)	50.6(2.2)	49.4(2.2)
GRADUATED H.S.	1994	2086	21.4(0.8) [4%]	251.6(1.2)	1.1(0.4)	19.9(1.6)	62.0(1.7)	38.0(1.7)
	1992	2137	23.7(0.8) [3%]	250.7(1.4)	1.1(0.3)	18.9(1.5)	60.6(1.8)	39.4(1.8)
EDUC AFTER H.S.	1994	1926	19.5(0.5) [3%]	265.5(1.3)	2.8(0.6)	33.0(1.5)	77.3(1.5)	22.7(1.5)
	1992	1806	19.1(0.5) [2%]	265.3(1.1)	2.8(0.7)	32.5(1.4)	76.3(1.5)	23.7(1.5)
GRAD COLLEGE	1994	4378	42.8(1.1) [2%]	269.5(0.9)	4.6(0.5)	39.9(1.2)	78.8(1.0)	21.2(1.0)
	1992	3946	41.3(1.2) [3%]	270.8(1.0)	5.0(0.7)	40.0(1.4)	79.9(1.0)	20.1(1.0)
UNKNOWN	1994	966	8.9(0.4) [4%]	238.3(1.6)	0.4(0.3)	11.8(1.3)	47.7(2.5)	52.3(2.5)
	1992	805	7.8(0.4) [6%]	237.5(2.0)	0.4(0.4)	12.0(2.0)	45.1(2.6)	54.9(2.6)
TYPE OF SCHOOL								
PUBLIC	1994	8327	89.4(1.0) [1%]	257.3(0.8)	2.4(0.3)	27.2(0.9)	67.3(0.9)	32.7(0.9)
	1992	7656	89.4(0.8) [1%]	257.5(1.0)	2.4(0.3)	26.9(1.1)	67.5(1.1)	32.5(1.1)
PRIVATE/CATHOLIC	1994	1808	10.6(1.0) [9%]	279.3(1.4)	6.2(0.8)	49.4(2.3)	88.5(1.3)	11.5(1.3)
	1992	1808	10.6(0.8) [8%]	278.3(2.0)	7.4(1.3)	48.4(3.2)	86.6(1.5)	13.4(1.5)
PRIVATE ONLY	1994	672	4.0(0.7) [17%]	280.4(2.4)	6.8(1.4)	50.1(3.9)	89.3(2.1)	10.7(2.1)
	1992	621	4.1(0.8) [19%]	283.1(3.0)	9.5(2.6)	54.2(4.4)	90.0(2.5)	10.0(2.5)
CATHOLIC ONLY	1994	1136	6.6(0.6) [10%]	278.6(1.3)	5.8(1.1)	49.0(2.1)	88.0(1.3)	12.0(1.3)
	1992	1187	6.5(0.6) [10%]	275.2(1.9)	6.0(1.0)	44.6(2.8)	84.5(1.6)	15.5(1.6)

BEST COPY AVAILABLE

Table J-11

NAEP 1992 and 1994 Reading Assessments, Grade 12 Weighted Percentages and Composite Proficiency Means, Percent of Students by Achievement Levels: Original Results

		N	WEIGHTED PCT [CV]	MEAN	ADVANCED	PRFCIENT	BASIC	< BASIC
-- TOTAL --	1994	9935	100.0(0.0) [1%]	286.5(0.7)<	3.5(0.4)	33.6(1.0)<	70.3(0.8)<	29.7(0.8)>
	1992	9856	100.0(0.0) [1%]	291.0(0.6)	3.3(0.3)	36.9(0.8)	75.2(0.7)	24.8(0.7)
SEX								
MALE	1994	4970	49.6(0.8) [2%]	279.6(0.8)<	1.7(0.3)	26.7(1.1)<	64.0(1.1)<	36.0(1.1)>
	1992	4713	48.6(0.6) [2%]	285.8(0.7)	2.1(0.3)	31.1(1.1)	70.3(1.1)	29.7(1.1)
FEMALE	1994	4965	50.4(0.8) [2%]	293.3(0.8)<	5.3(0.7)	40.3(1.2)	76.4(1.1)	23.6(1.1)
	1992	5143	51.4(0.6) [1%]	296.0(0.7)	4.4(0.4)	42.4(1.2)	79.8(0.9)	20.2(0.9)
REGION								
NORTHEAST	1994	2289	20.3(0.5) [3%]	287.6(1.7)<	4.1(0.9)	34.1(1.9)	71.3(2.0)	28.7(2.0)
	1992	2288	24.0(0.6) [3%]	293.0(1.2)	3.9(0.5)	40.2(1.6)	76.4(1.6)	23.6(1.6)
SOUTHEAST	1994	2777	23.4(0.7) [3%]	281.2(1.2)	2.5(0.5)	27.3(1.9)	64.6(1.3)	35.4(1.3)
	1992	2714	23.1(0.6) [3%]	283.5(1.1)	2.0(0.3)	28.0(1.4)	67.7(1.4)	32.3(1.4)
CENTRAL	1994	2005	27.0(0.7) [3%]	290.0(1.2)<	3.8(0.6)	36.9(1.9)	74.3(1.7)	25.7(1.7)
	1992	2339	26.1(0.6) [2%]	294.4(1.1)	3.4(0.4)	40.4(1.6)	79.1(1.4)	20.9(1.4)
WEST	1994	2864	29.4(0.8) [3%]	286.7(1.4)<	3.7(0.9)	35.1(1.9)	70.4(1.3)	29.6(1.3)
	1992	2515	26.8(0.8) [3%]	292.5(1.6)	3.7(0.6)	38.3(2.2)	76.7(2.0)	23.3(2.0)
RACE/ETHNICITY								
WHITE	1994	6510	73.3(0.3) [1%]	293.2(0.7)<	4.4(0.5)	39.7(1.1)	77.2(0.9)<	22.8(0.9)>
	1992	6728	71.6(0.4) [1%]	297.0(0.6)	4.1(0.3)	43.3(0.9)	81.7(0.8)	18.3(0.8)
BLACK	1994	1548	12.7(0.3) [3%]	263.9(1.6)<	0.4(0.2)	11.7(1.2)	46.1(2.2)	53.9(2.2)
	1992	1603	15.0(0.4) [3%]	271.8(1.5)	0.4(0.2)	15.6(1.5)	53.8(2.5)	46.2(2.5)
HISPANIC	1994	1175	8.4(0.3) [4%]	268.5(1.5)<	1.1(0.5)	18.0(1.7)	52.2(2.5)	47.8(2.5)
	1992	1005	8.8(0.4) [5%]	277.2(2.4)	1.4(0.7)	21.0(2.8)	61.2(3.2)	38.8(3.2)
ASIAN	1994	385	2.9(0.3) [10%]	278.8(2.8)	2.7(1.2)	30.2(2.9)	63.2(3.0)	36.8(3.0)
PACIFIC ISLANDER	1994	172	1.2(0.3) [21%]!	279.7(3.9)	2.3(1.3)	24.7(5.9)	66.4(5.4)	33.6(5.4)
AMERICAN INDIAN	1994	103	1.1(0.4) [34%]!	273.5(5.4)	2.0(2.8)	17.8(6.8)	55.1(5.9)	44.9(5.9)
	1992	59	0.5(0.1) [20%]	271.8(5.3)$	1.4(1.2)$	23.6(6.9)$	52.4(7.7)$	47.6(7.7)$

! INTERPRET WITH CAUTION: SAMPLING ERROR CANNOT BE ESTIMATED ACCURATELY SINCE COEFFICIENT OF VARIATION OF ESTIMATED NUMBER OF STUDENTS EXCEEDS 20%

$ SAMPLE SIZE IS INSUFFICIENT TO PERMIT A RELIABLE ESTIMATE

> INDICATES A SIGNIFICANT INCREASE (OR DECREASE "<") BETWEEN 1992 AND 1994

BEST COPY AVAILABLE

Table J-11 (continued)

NAEP 1992 and 1994 Reading Assessments, Grade 12 Weighted Percentages and Composite Proficiency Means, Percent of Students by Achievement Levels: Original Results

		N	WEIGHTED PCT [CV]	MEAN	ADVANCED	PRFCIENT	BASIC	< BASIC
PARENTS' EDUCATION LEVEL								
LESS THAN H.S.	1994	835	7.4(0.4) [6%]	264.7(1.5)<	0.5(0.3)	13.5(1.5)	46.6(2.7)	53.4(2.7)
	1992	771	7.5(0.4) [6%]	273.5(1.5)	0.4(0.3)	17.9(2.1)	56.5(2.6)	43.5(2.6)
GRADUATED H.S.	1994	1976	20.7(0.7) [4%]	275.8(1.3)<	1.3(0.4)	21.4(1.7)	60.5(1.8)	39.5(1.8)
	1992	2063	21.9(0.5) [3%]	281.4(0.8)	1.2(0.3)	24.7(1.2)	66.1(1.6)	33.9(1.6)
EDUC AFTER H.S.	1994	2447	25.4(0.7) [3%]	288.2(1.0)<	2.9(0.5)	33.0(1.4)	73.2(1.5)	26.8(1.5)
	1992	2551	26.7(0.6) [2%]	292.7(0.8)	2.9(0.5)	37.8(1.2)	78.0(1.0)	22.0(1.0)
GRAD COLLEGE	1994	4348	43.5(1.0) [2%]	297.2(1.0)	5.7(0.8)	45.1(1.4)	80.3(0.8)<	19.7(0.8)>
	1992	4190	41.3(0.9) [2%]	300.5(0.8)	5.3(0.6)	48.1(1.1)	84.1(0.8)	15.9(0.8)
UNKNOWN	1994	279	2.5(0.2) [8%]	246.6(2.8)<	0.2(0.2)	5.3(1.6)	27.0(3.3)	73.0(3.3)
	1992	251	2.4(0.2) [7%]	256.9(2.8)	0.3(0.7)	7.7(2.2)	38.3(6.1)	61.7(6.1)
TYPE OF SCHOOL								
PUBLIC	1994	8044	89.3(1.1) [2%]	284.9(0.7)<	3.2(0.4)	31.9(1.0)	68.8(0.8)<	31.2(0.8)>
	1992	7766	86.9(1.2) [2%]	288.7(0.7)	2.6(0.3)	34.1(0.9)	73.0(0.9)	27.0(0.9)
PRIVATE/CATHOLIC	1994	1875	10.3(1.0) [10%]	300.7(1.9)<	6.7(0.8)	48.5(2.6)	83.6(1.8)<	16.4(1.8)>
	1992	2090	13.1(1.2) [9%]	306.7(1.3)	7.5(0.7)	55.7(2.4)	89.7(1.1)	10.3(1.1)
PRIVATE ONLY	1994	711	4.0(0.6) [14%]	306.2(2.2)	9.5(1.4)	55.9(3.0)	86.9(1.9)	13.1(1.9)
	1992	643	4.2(0.7) [17%]	307.7(3.0)	10.2(1.5)	57.8(4.3)	87.1(2.6)	12.9(2.6)
CATHOLIC ONLY	1994	1164	6.3(0.9) [15%]	297.2(2.4)<	4.9(1.1)	43.9(3.4)	81.5(2.3)<	18.5(2.3)>
	1992	1447	8.8(1.2) [14%]	306.3(1.5)	6.2(0.8)	54.7(2.8)	90.9(1.2)	9.1(1.2)

! INTERPRET WITH CAUTION: SAMPLING ERROR CANNOT BE ESTIMATED ACCURATELY SINCE COEFFICIENT OF VARIATION OF ESTIMATED NUMBER OF STUDENTS EXCEEDS 20%

$ SAMPLE SIZE IS INSUFFICIENT TO PERMIT A RELIABLE ESTIMATE

> INDICATES A SIGNIFICANT INCREASE (OR DECREASE "<") BETWEEN 1992 AND 1994

960 961

Table J-12

NAEP 1992 and 1994 Reading Assessments
Grade 12 Weighted Percentages and Composite Proficiency Means
Percent of Students by Achievement Levels: Revised Results

		N	WEIGHTED PCT [CV]	MEAN	ADVANCED	PRFCIENT	BASIC	< BASIC
-- TOTAL --	1994	9935	100.0(0.0) [1%]	287.3(0.7)<	4.2(0.5)	36.3(1.0)<	74.5(0.7)<	25.5(0.7)>
	1992	9856	100.0(0.0) [1%]	292.1(0.6)	3.9(0.3)	40.2(0.8)	79.7(0.6)	20.3(0.6)
SEX								
MALE	1994	4970	49.6(0.8) [2%]	280.4(0.8)<	2.2(0.3)	29.3(1.1)<	68.5(1.1)<	31.5(1.1)>
	1992	4713	48.6(0.6) [2%]	286.9(0.7)	2.5(0.4)	34.1(1.1)	75.2(1.0)	24.8(1.0)
FEMALE	1994	4965	50.4(0.8) [2%]	294.2(0.8)<	6.2(0.9)	43.2(1.1)	80.4(1.0)<	19.6(1.0)>
	1992	5143	51.4(0.6) [1%]	297.1(0.7)	5.3(0.4)	46.0(1.3)	83.9(0.7)	16.1(0.7)
REGION								
NORTHEAST	1994	2289	20.3(0.5) [3%]	288.4(1.7)<	4.8(1.0)	36.8(1.9)	75.5(1.7)	24.5(1.7)
	1992	2288	24.0(0.6) [3%]	294.1(1.1)	4.7(0.6)	43.5(1.7)	80.5(1.5)	19.5(1.5)
SOUTHEAST	1994	2777	23.4(0.7) [3%]	282.1(1.2)	3.0(0.6)	29.6(2.0)	69.8(1.2)	30.2(1.2)
	1992	2714	23.1(0.6) [3%]	284.7(1.1)	2.4(0.4)	31.0(1.4)	72.9(1.4)	27.1(1.4)
CENTRAL	1994	2005	27.0(0.7) [3%]	290.8(1.2)<	4.7(0.6)	40.1(1.6)	78.0(1.5)<	22.0(1.5)>
	1992	2339	26.1(0.6) [2%]	295.5(1.1)	4.2(0.5)	43.9(1.7)	83.5(1.1)	16.5(1.1)
WEST	1994	2864	29.4(0.8) [3%]	287.6(1.4)<	4.3(1.1)	37.8(1.9)	74.3(1.3)<	25.7(1.3)>
	1992	2515	26.8(0.8) [3%]	293.6(1.5)	4.4(0.6)	41.7(2.5)	81.0(1.5)	19.0(1.5)
RACE/ETHNICITY								
WHITE	1994	6510	73.3(0.3) [1%]	294.0(0.7)<	5.3(0.7)	42.8(1.1)	81.0(0.7)<	19.0(0.7)>
	1992	6728	71.6(0.4) [1%]	298.0(0.6)	4.9(0.4)	47.0(1.0)	85.6(0.7)	14.4(0.7)
BLACK	1994	1548	12.7(0.3) [3%]	264.9(1.6)<	0.6(0.2)	13.3(1.5)	52.0(2.2)	48.0(2.2)
	1992	1603	15.0(0.4) [3%]	273.2(1.4)	0.5(0.3)	17.7(1.5)	60.6(2.3)	39.4(2.3)
HISPANIC	1994	1175	8.4(0.3) [4%]	269.6(1.5)<	1.2(0.5)	19.8(1.8)	57.7(2.4)	42.3(2.4)
	1992	1005	8.8(0.4) [5%]	278.4(2.3)	1.5(0.7)	23.8(3.2)	66.4(2.5)	33.6(2.5)
ASIAN	1994	385	2.9(0.3) [10%]	279.9(2.8)	3.5(1.6)	32.5(3.0)	66.5(3.1)	33.5(3.1)
PACIFIC ISLANDER	1994	172	1.2(0.3) [21%]!	280.4(3.9)	2.6(1.5)	27.0(5.0)	70.5(4.3)	29.5(4.3)
AMERICAN INDIAN	1994	103	1.1(0.4) [34%]!	274.6(5.3)	2.1(2.8)	20.1(6.7)	60.8(6.5)	39.2(6.5)
	1992	59	0.5(0.1) [20%]	273.3(5.2)$	1.9(1.5)$	25.8(7.3)$	57.4(8.7)$	42.6(8.7)$

! INTERPRET WITH CAUTION: SAMPLING ERROR CANNOT BE ESTIMATED ACCURATELY SINCE COEFFICIENT OF VARIATION OF ESTIMATED NUMBER OF STUDENTS EXCEEDS 20%

$ SAMPLE SIZE IS INSUFFICIENT TO PERMIT A RELIABLE ESTIMATE

> INDICATES A SIGNIFICANT INCREASE (OR DECREASE "<") BETWEEN 1992 AND 1994

BEST COPY AVAILABLE

Table J-12 (continued)

NAEP 1992 and 1994 Reading Assessments
Grade 12 Weighted Percentages and Composite Proficiency Means
Percent of Students by Achievement Levels: Revised Results

		N	WEIGHTED PCT [CV]	MEAN	ADVANCED	PRFCIENT	BASIC	< BASIC
PARENTS' EDUCATION LEVEL								
LESS THAN H.S.	1994	835	7.4(0.4) [6%]	265.8(1.5)<	0.7(0.4)	15.3(1.5)	52.5(2.5)<	47.5(2.5)>
	1992	771	7.5(0.4) [6%]	274.9(1.4)	0.4(0.3)	20.8(2.1)	62.6(2.2)	37.4(2.2)
GRADUATED H.S.	1994	1976	20.7(0.7) [4%]	276.7(1.3)<	1.7(0.5)	23.8(1.7)	65.7(1.7)<	34.3(1.7)>
	1992	2063	21.9(0.5) [3%]	282.7(0.8)	1.5(0.3)	27.7(1.4)	71.9(1.2)	28.1(1.2)
EDUC AFTER H.S.	1994	2447	25.4(0.7) [3%]	289.0(1.0)<	3.5(0.5)	36.0(1.4)	77.6(1.3)<	22.4(1.3)>
	1992	2551	26.7(0.6) [2%]	293.8(0.8)	3.5(0.5)	41.4(1.4)	82.7(1.0)	17.3(1.0)
GRAD COLLEGE	1994	4348	43.5(1.0) [2%]	298.0(1.0)<	6.7(1.1)	48.1(1.4)	83.7(0.7)<	16.3(0.7)>
	1992	4190	41.3(0.9) [2%]	301.4(0.8)	6.4(0.6)	51.6(1.3)	87.4(0.7)	12.6(0.7)
UNKNOWN	1994	279	2.5(0.2) [8%]	247.9(2.7)<	0.4(0.7)	6.3(2.3)	31.7(3.3)	68.3(3.3)
	1992	251	2.4(0.2) [7%]	258.4(2.8)	0.3(0.7)	9.7(2.0)	44.0(4.9)	56.0(4.9)
TYPE OF SCHOOL								
PUBLIC	1994	8044	89.3(1.1) [2%]	285.8(0.7)<	3.8(0.5)	34.6(1.0)	73.2(0.7)<	26.8(0.7)>
	1992	7766	86.9(1.2) [2%]	289.8(0.7)	3.2(0.3)	37.3(0.9)	77.8(0.7)	22.2(0.7)
PRIVATE/CATHOLIC	1994	1875	10.3(1.0) [10%]	301.4(1.9)<	7.8(0.9)	51.8(2.7)	86.5(1.7)<	13.5(1.7)>
	1992	2090	13.1(1.2) [9%]	307.6(1.3)	9.1(0.7)	59.6(2.2)	92.0(0.8)	8.0(0.8)
PRIVATE ONLY	1994	711	4.0(0.6) [14%]	306.9(2.2)	10.7(1.5)	59.0(3.1)	89.2(2.0)	10.8(2.0)
	1992	643	4.2(0.7) [17%]	308.5(2.9)	12.3(1.8)	61.2(3.8)	89.2(2.2)	10.8(2.2)
CATHOLIC ONLY	1994	1164	6.3(0.9) [15%]	297.9(2.4)<	5.9(1.1)	47.3(3.7)	84.8(2.2)<	15.2(2.2)>
	1992	1447	8.8(1.2) [14%]	307.1(1.5)	7.5(0.7)	58.8(2.6)	93.4(0.9)	6.6(0.9)

BEST COPY AVAILABLE

Appendix K

THE INFORMATION WEIGHTING ERROR

Appendix K

THE INFORMATION WEIGHTING ERROR[1]

Susan C. Loomis, Luz Bay, and Wen-Hung Chen
American College Testing

The Error

In the process of recomputing the reading cutscores set in 1992 for the three achievement levels, an error in the information weighting function was detected. The error affected data for all achievement levels set in 1992: reading and mathematics. The Muraki information weighting function published in 1993 was used in the 1994 programs to compute achievement levels, so only 1992 levels are affected.

The procedures used for 1992 were printed and reported in numerous places. No one had detected an error. The psychometrician who developed the programs for the 1994 process used Muraki's information weighting function because he found it to be more straightforward than the 1992 procedure.

The 1992 equation[2] is as follows:

$$I_j(\theta) = D^2 a_j^2 \sum_{c=1}^{m_j} P_{jc}(\theta)[1 - P_{jc}(\theta)] \; . \qquad (1)$$

The 1994 equation[3] is as follows:

$$I_j(\theta) = D^2 a_j^2 \sum_{c=1}^{m_j} [T_c - \bar{T}_j(\theta)]^2 P_{jc}(\theta) \; , \qquad (2)$$

where

$$\bar{T}_j(\theta) = \sum_{c=1}^{m_j} T_c P_{jc}(\theta) \; .$$

---

[1] This appendix was reviewed by Mark Reckase and Alan Nicewander of American College Testing.

[2] Luecht, Richard M. (April, 1993). *Using IRT to improve the standard setting process for dichotomous and polytomous items.* Paper presented at the annual meeting of the National Council on Measurement in Education, Atlanta, Georgia. Notations for this equation were modified to correspond to those of Equation 2. The reader will need to refer to the articles for a complete explanation of the equations.

[3] Muraki, Eiji (1993). Information functions of the generalized partial credit model. *Applied Psychological Measurement, 17(4)*, 351-362.

Analysis of the Error: Magnitude

The differences in achievement levels reported for 1992 and 1994 and the corrected achievement levels are due both to the error in item parameters and to the error in information weights. The cutscores and percentages of students scoring at or above each for each achievement level are reported in Tables K-1 and K-2. Data in Table K-1 are the previously reported (incorrect) data, and data in Table K-2 are the corrected data.

The maximum difference in cutscores originally reported and the corrected cutscores is found for grade 4 at the Advanced level:

$$((\text{original cutscore} = 275) - (\text{corrected cutscore} = 268)) = 7 \text{ points}.$$

The differences attributable to each error appear to be rather small in most cases.

Table K-3 reports the differences in cutscores due to the two errors, examined one at a time. Relative to the *correct* data, the information weighting error generally resulted in a higher composite cutscore, while the recoding error resulting in incorrect item parameters generally resulted in a lower composite cutscore.

Figures K-1, K-2, and K-3 show comparisons of percentages of students who scored at or above each achievement level in 1994.[4] The center bar on each of these graphs shows the percentage of students who scored at or above the achievement level in 1994 using both correct item parameters and correct information weights. The bar on the left shows the percentage of students who would have scored at or above the achievement level in 1994 computed with the *correct* item parameters and *incorrect* information weights. The bar on the right shows the percentage computed with the *incorrect* item parameters and the *correct* information weights. These graphs show that the effect of each error was about the same, with respect to the distribution of student scores relative to the cutscores for achievement levels. The greatest difference due to the error in item parameters only is seen for grade 12 at the Advanced level. The greatest difference due to the error in information weights only is seen at the Advanced level for both grades 4 and 8.

---

[4] Distribution data were not recomputed for the 1992 data using incorrect item parameters and corrected information weights, so these comparisons cannot be presented for 1994.

Table K-1
Reading Cutpoints and Percents At or Above as Reported

Grade		Basic	Proficient	Advanced
	Cutpoint	212	243	275
4	%≥92 Dist	59.0	25.3	4.5
	%≥94 Dist	57.7	25.4	4.6
	Cutpoint	244	283	328
8	%≥92 Dist	68.8	27.5	2.1
	%≥94 Dist	69.1	27.8	1.9
	Cutpoint	269	304	348
12	%≥92 Dist	75.2	37.0	3.2
	%≥94 Dist	70.3	33.6	3.5

Table K-2
Corrected Reading Cutpoints and Percents At or Above

Grade		Basic	Proficient	Advanced
	Cutpoint	208	238	268
4	%≥92 Dist	62.1	28.6	6.4
	%≥94 Dist	60.5	29.6	7.4
	Cutpoint	243	281	323
8	%≥92 Dist	69.5	29.2	2.9
	%≥94 Dist	69.4	29.5	2.8
	Cutpoint	265	302	346
12	%≥92 Dist	79.7	40.2	3.9
	%≥94 Dist	74.5	36.3	4.2

Table K-3

Composite NAEP Scale Cutpoint Differences in Reading
Due to Errors

Achievement Level Cutpoint	Information Weighting*	Item Parameters**
Grade 4		
Basic	3	5
Proficient	3	2
Advanced	5	-1
Grade 8		
Basic	1	1
Proficient	1	-1
Advanced	5	-4
Grade 12		
Basic	0	-1
Proficient	1	-3
Advanced	2	-9

**Note: Difference = Incorrect - Correct, based on correct item parameters. If the recoding of data had been correct, the cutpoints would have been in error by these amounts, due to the incorrect information weighting function.*

***Note: Difference = Incorrect - Correct, based on correct information weights. If the correct information weighting function had been used, the cutpoints would have been in error by these amounts due to the recoding error resulting in incorrect item parameters.*

Figure K-1
NAEP Reading Achievement Levels:
Cutpoints and 1994 Distribution Data
Grade 4

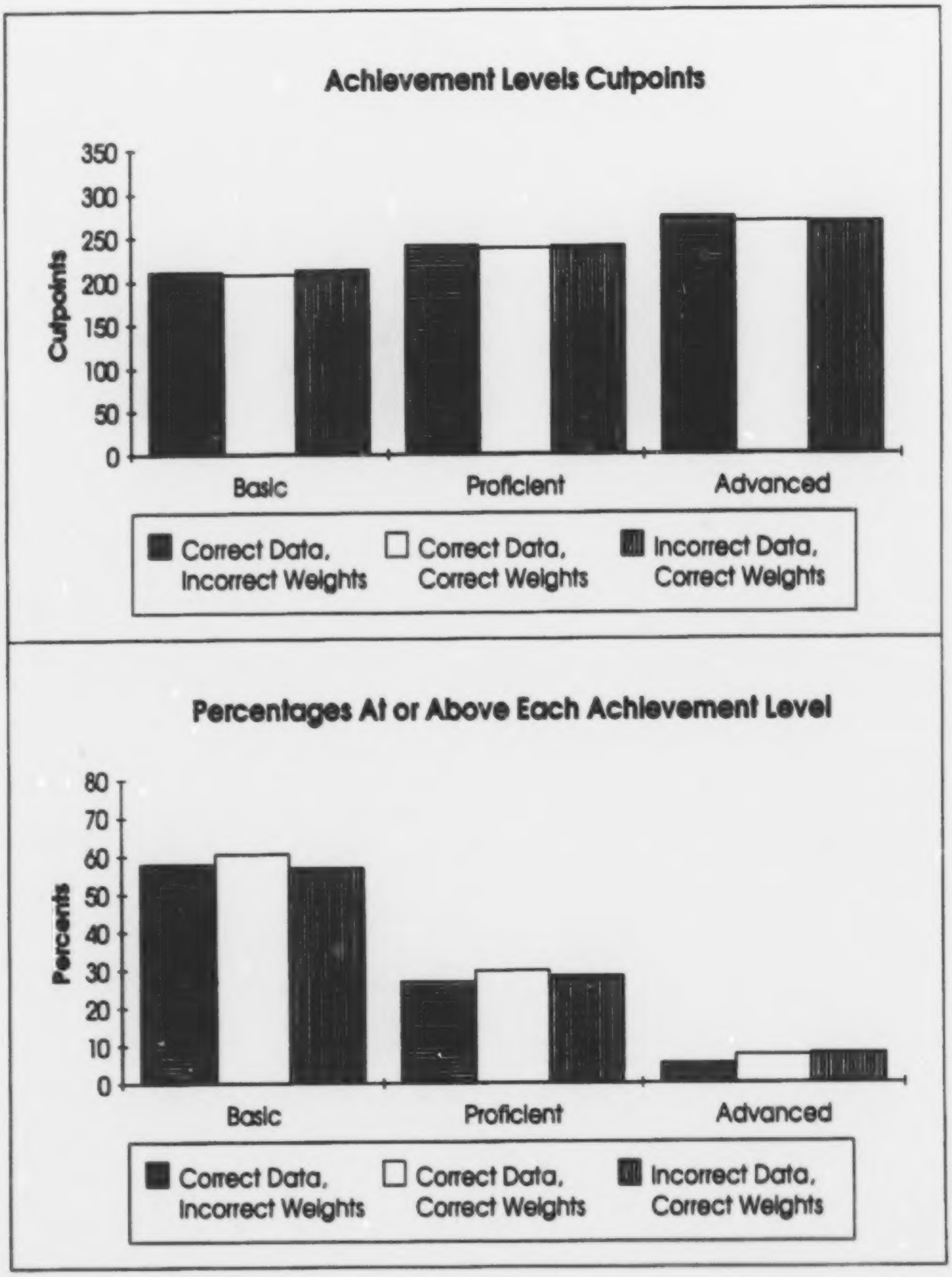


Figure K-2

NAEP Reading Achievement Levels:
Cutpoints and 1994 Distribution Data
Grade 8

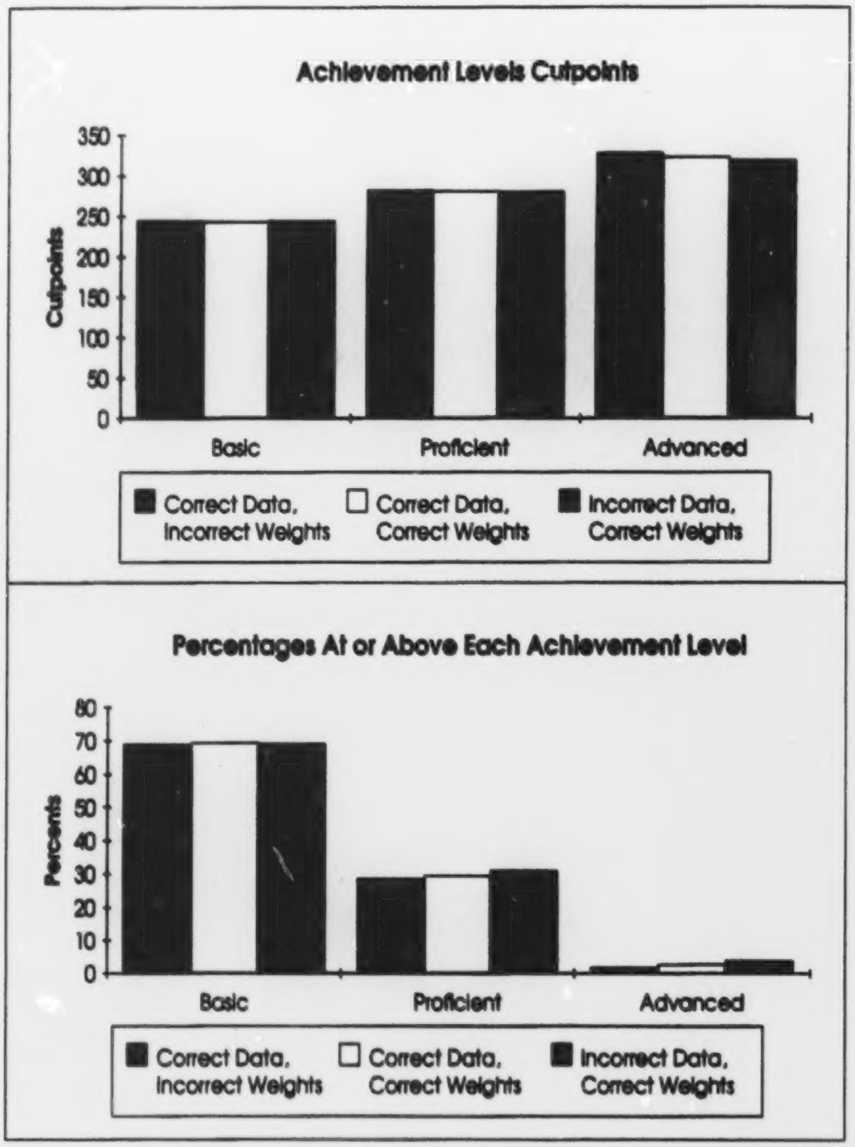
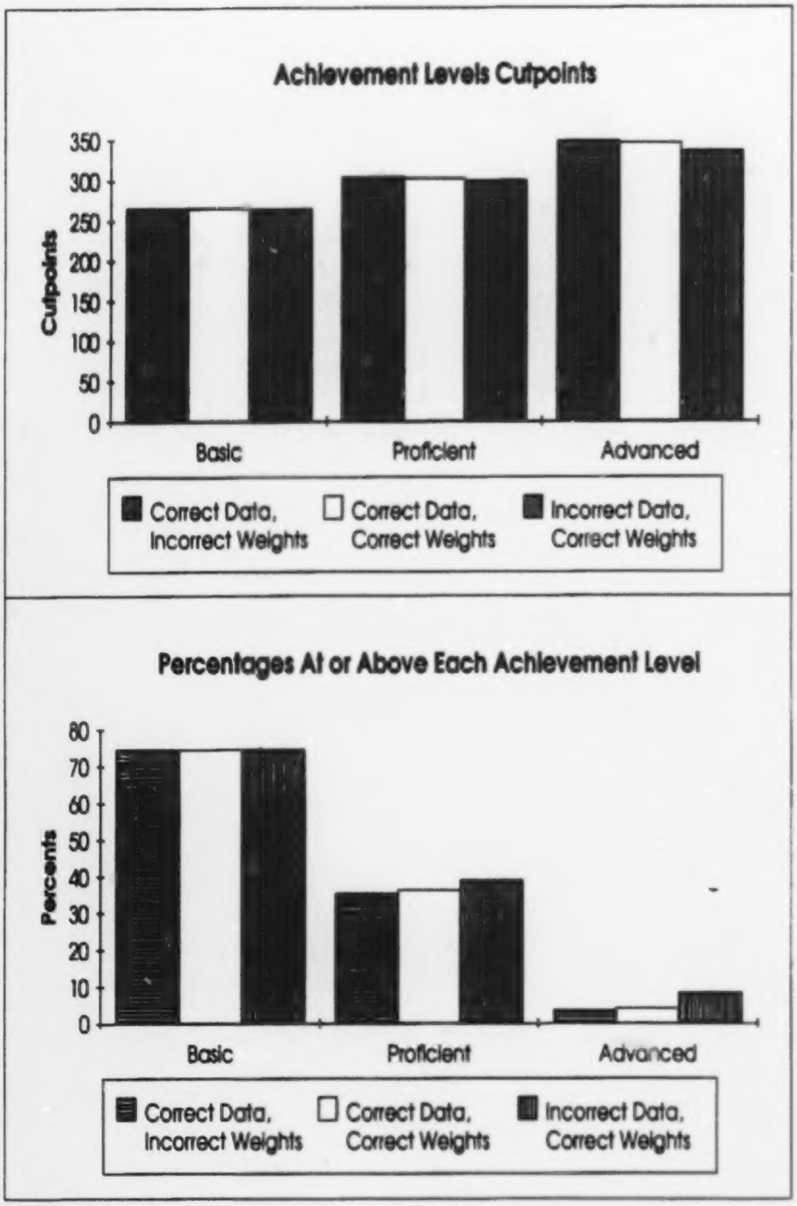


Figure K-3
NAEP Reading Achievement Levels:
Cutpoints and 1994 Distribution Data
Grade 12



Analysis of the Error: The Information Weighting Functions

Various analyses were conducted to determine what, if any, general conclusions could be drawn to help inform users of NAEP achievement levels data about the factors related to differences in cutscores due to the information weighting error.

Item ratings are collected from two groups of panelists at each grade level. These groups are called item rating groups, and panelists are assigned to an item rating group so that the two are as equivalent as possible in terms of panelist type (teacher, career educator, or general public; gender; race/ethnicity; and region of residence). These item rating groups rate slightly over half of all items at their grade level. Item rating pools are developed so that the items in each are as equivalent as possible in terms of item difficulty, item format (multiple choice, short constructed response, and extended constructed response), test time for the block, and so forth. Item blocks remain intact for the item rating pools. At least one block (a "common block") is rated by all panel members (i.e., both item rating groups, in the grade group).

Item ratings are placed on the NAEP scale by computing a theta value for the dichotomous items and for the polytomous items in each scale for each rating group. Information weights are applied for the polytomous items at the scale level before computing the scale score for both dichotomous and polytomous items.

Information weighting functions were graphed for the polytomous items in each scale. Scales contain items from various blocks. Some scales contain few polytomous items. The number of items included in the graphs is further reduced by the fact that scale rating data by item type (polytomous or dichotomous) is by item rating group (A or B). Thus, some scale scores for a rating group are based on ratings for only one or two polytomous items.

Figures K-4 through K-7 show the "old," (i.e., incorrect, information weighting function used for computing the cutscores in 1992) and the "new," (i.e., correct, information weighting function used for computing cutscores for grade 4). The graphs in Figure K-4 show information weighting functions for which the greatest differences in the two (old and new) were found (grade 4, group A, scale 1, at the Advanced level). Figures K-8 through K-11 are graphs comparing the information weighting functions for some scales for grades 8 and 12. Note in Figures K-9 and K-10, there are negative differences (i.e., the "new" weights exceed the "old" weights) in the areas where information is maximum.

Table K-4 presents the information weights computed for each rating group and each scale for the reading NAEP achievement levels. Those data show that there is no *consistent* pattern of error caused by the incorrect information function. The corrected cutscores are consistently neither higher nor lower as a result of this error, but the impact of the error is generally to estimate a higher cutscore for the polytomous items.

Analysis of the Effect of Item Discrimination. Figures K-12 through K-15 show graphs for the correct ("new") and the incorrect ("old") information functions holding other parameters constant while varying the item discrimination parameter in the generalized partial credit IRT model. In general, the differences between correct and incorrect weights increase as item discrimination increases.

Figure K-4

Comparison Between the "Old" and "New" Information Functions for 1992 Reading, Grade 4, Group A, Scale 1

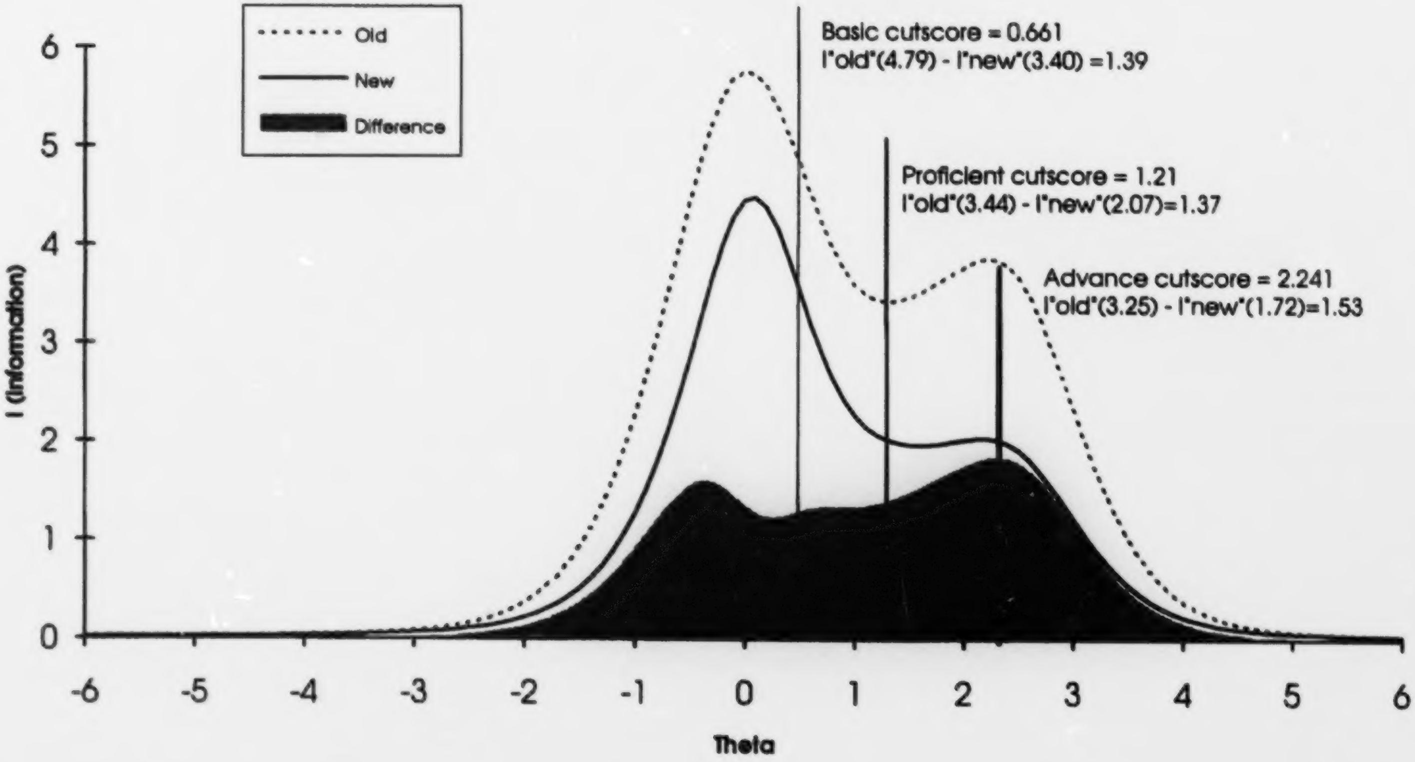
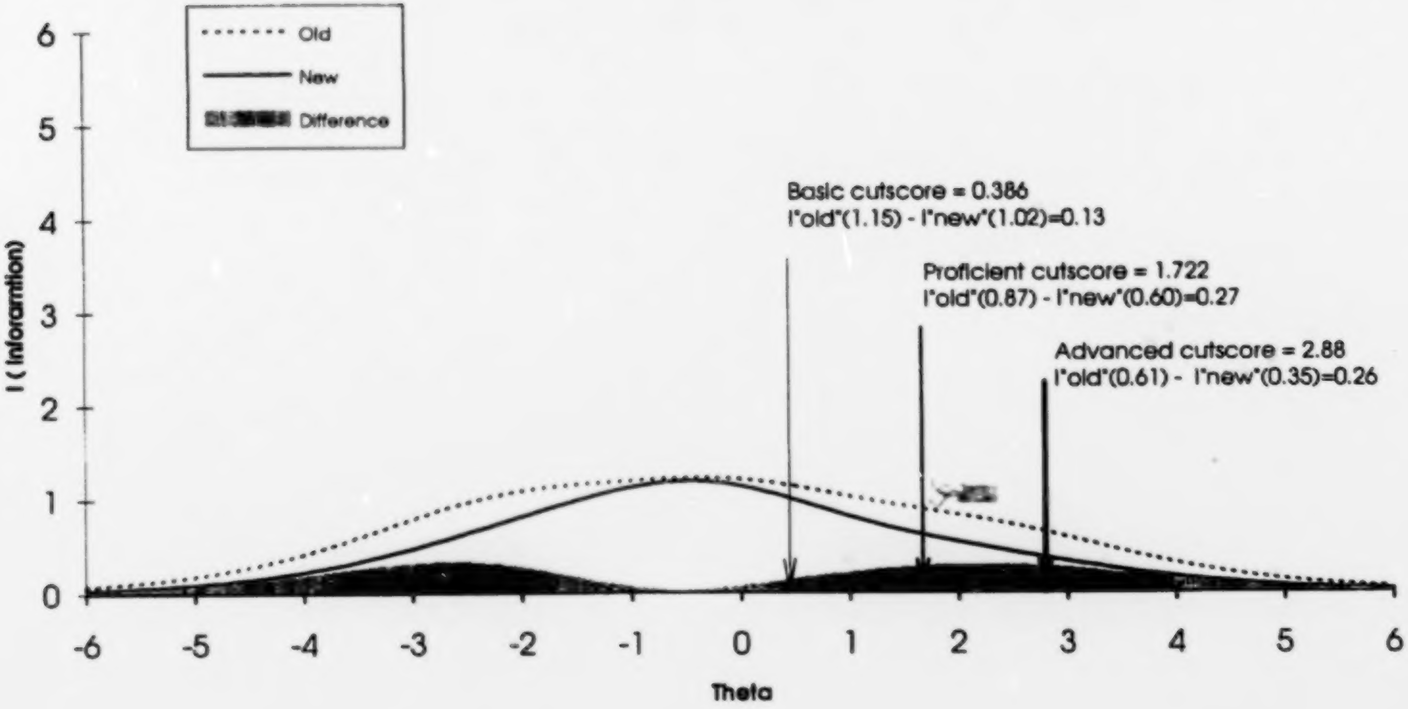


Figure K-5

Comparison Between the "Old" and "New" Information Functions for 1992 Reading, Grade 4, Group A, Scale 2



Basic cutscore = 0.386
I"old"(1.15) - I"new"(1.02)=0.13

Proficient cutscore = 1.722
I"old"(0.87) - I"new"(0.60)=0.27

Advanced cutscore = 2.88
I"old"(0.61) - I"new"(0.35)=0.26

Figure K-6
Comparison Between the "Old" and "New" Information Functions
for 1992 Reading, Grade 4, Group B, Scale 1

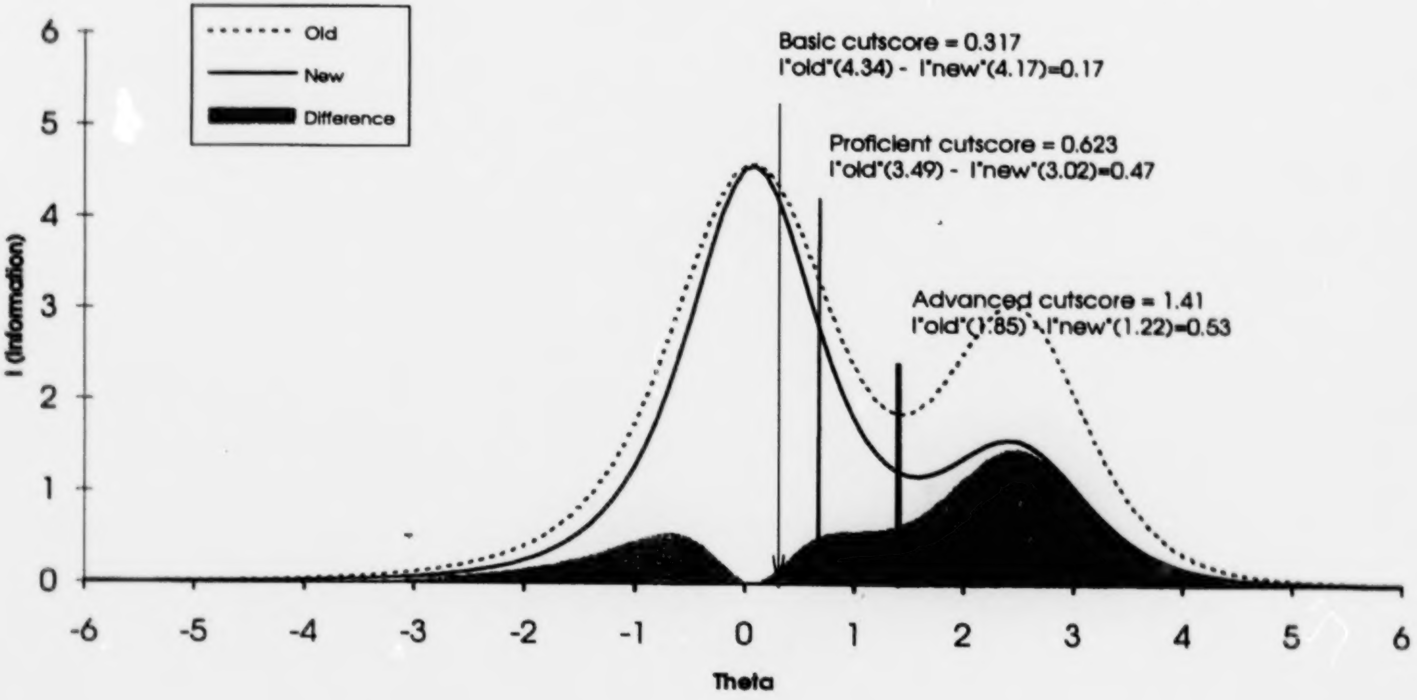
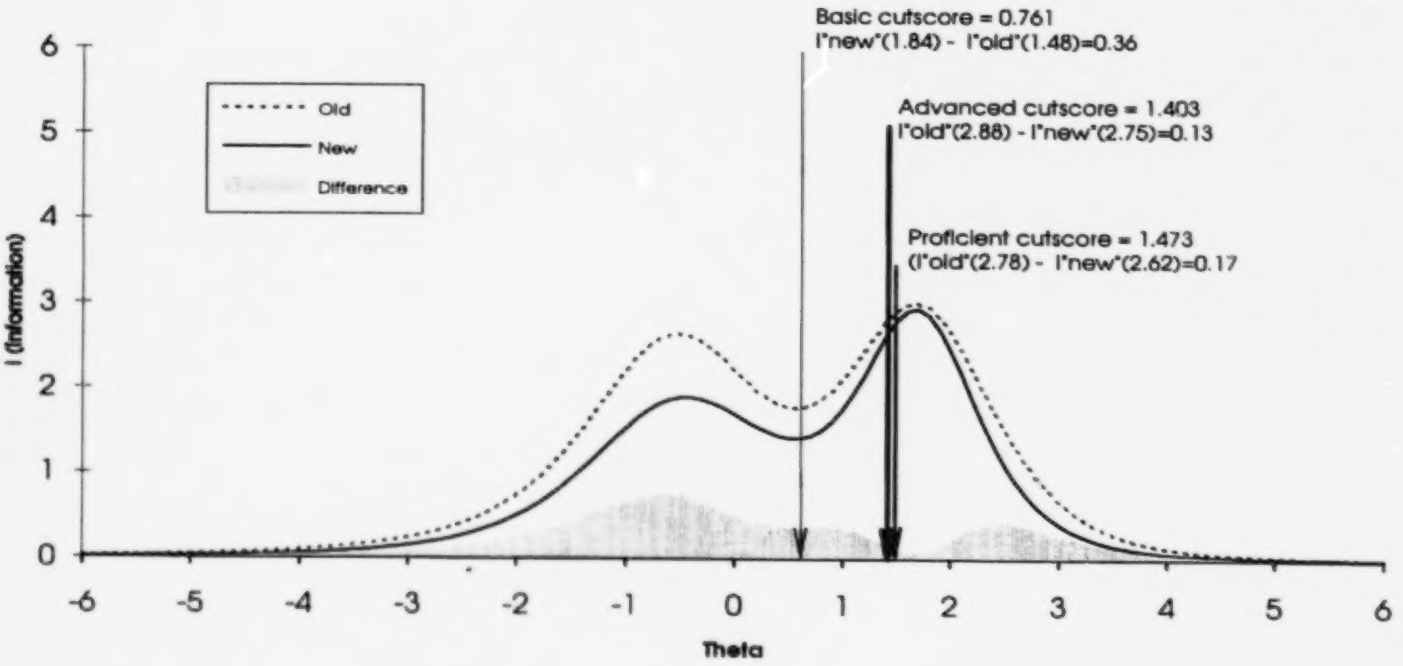


Figure K-7

Comparison Between the "Old" and "New" Information Functions for 1992 Reading, Grade 4, Group B, Scale 2



Basic cutscore = 0.761
I"new"(1.84) - I"old"(1.48)=0.36

Advanced cutscore = 1.403
I"old"(2.88) - I"new"(2.75)=0.13

Proficient cutscore = 1.473
(I"old"(2.78) - I"new"(2.62)=0.17

Figure K-8

Comparison Between the "Old" and "New" Information Functions for 1992 Reading, Grade 8, Group A, Scale 2

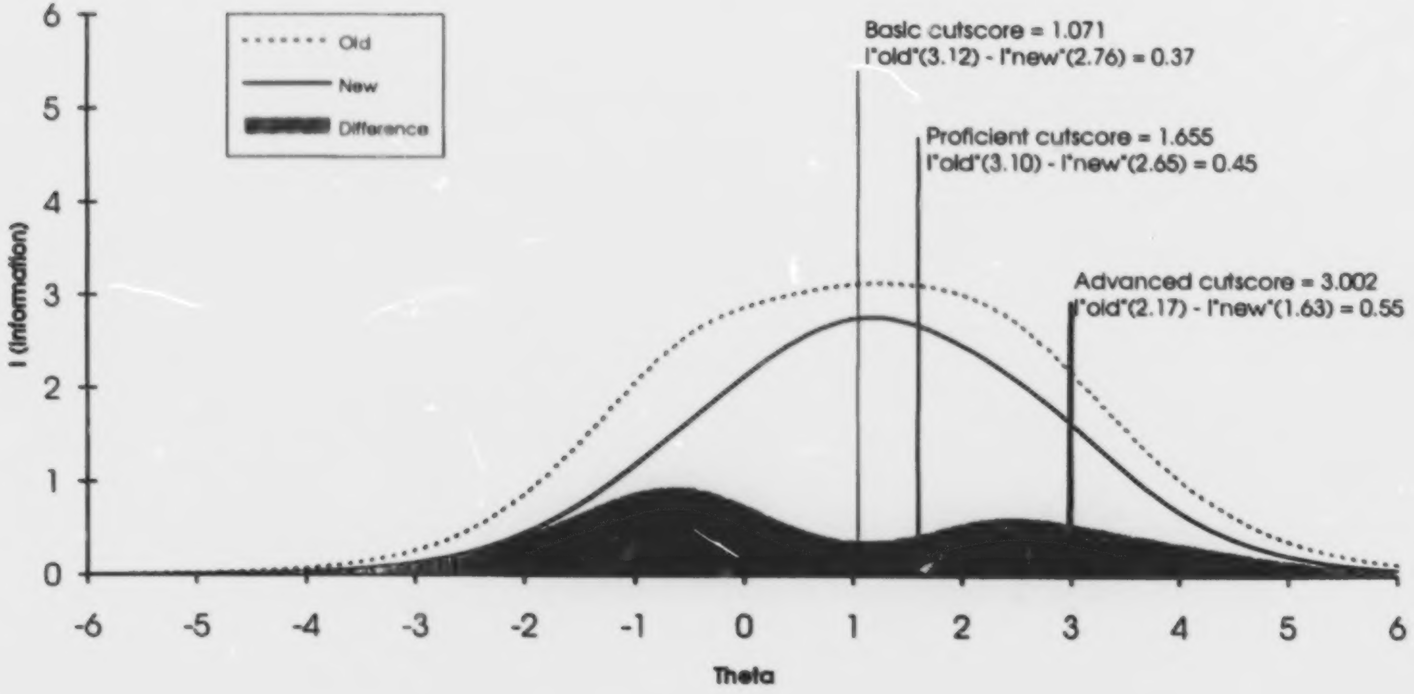
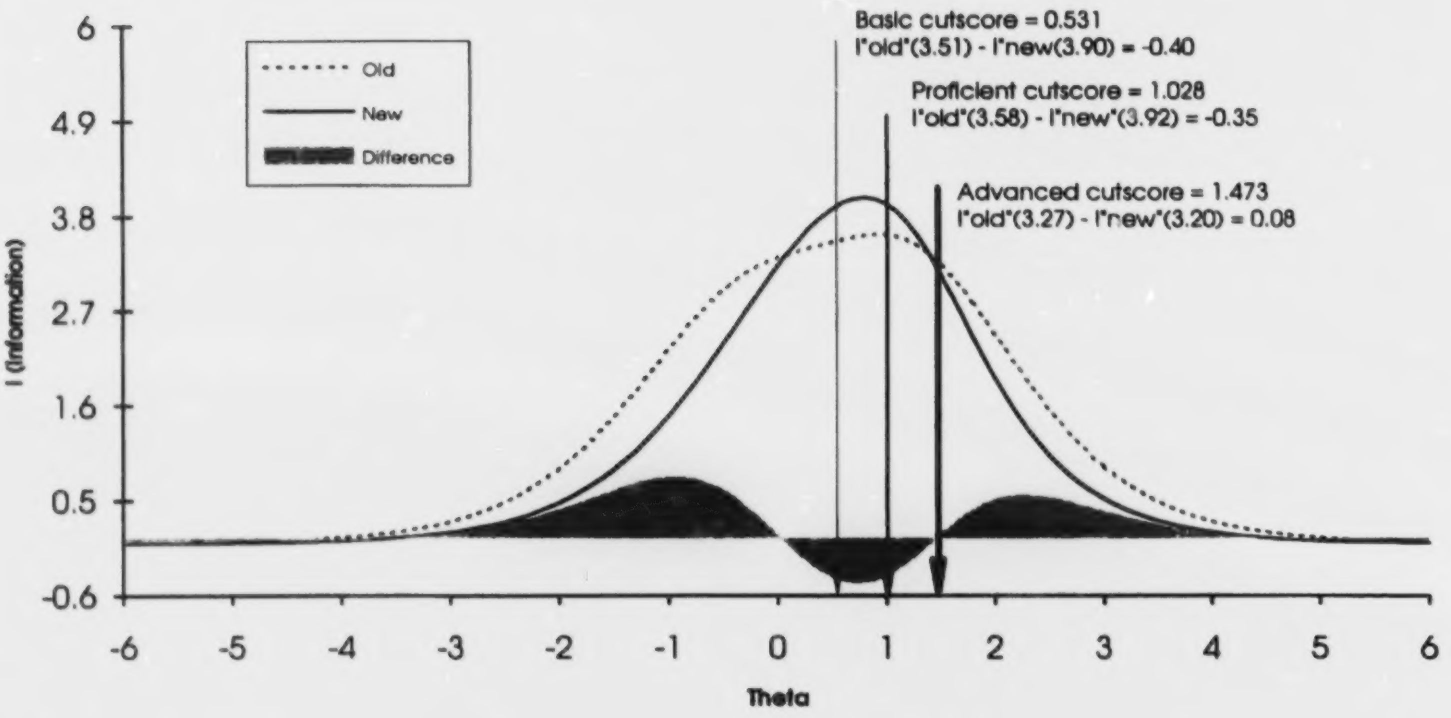


Figure K-9

Comparison Between the "Old" and "New" Information Functions for 1992 Reading, Grade 8, Group B, Scale 2



Basic cutscore = 0.531
I"old"(3.51) - I"new(3.90) = -0.40

Proficient cutscore = 1.028
I"old"(3.58) - I"new"(3.92) = -0.35

Advanced cutscore = 1.473
I"old"(3.27) - I"new"(3.20) = 0.08

Figure K-10

Comparison Between the "Old" and "New" Information Functions for 1992 Reading, Grade 12, Group A, Scale 2

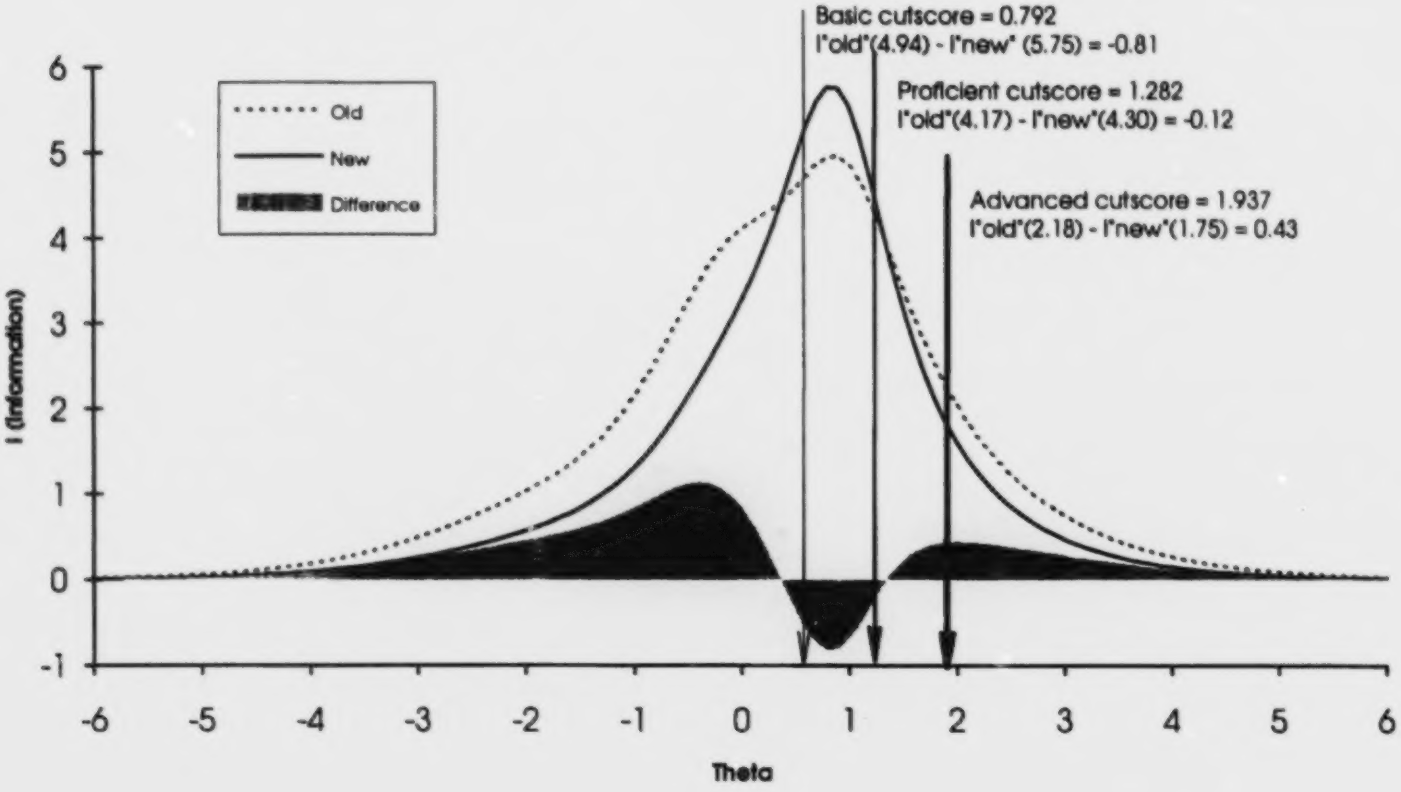


Figure K-11

Comparison Between the "Old" and "New" Information Functions
for 1992 Reading, Grade 12, Group B, Scale 3

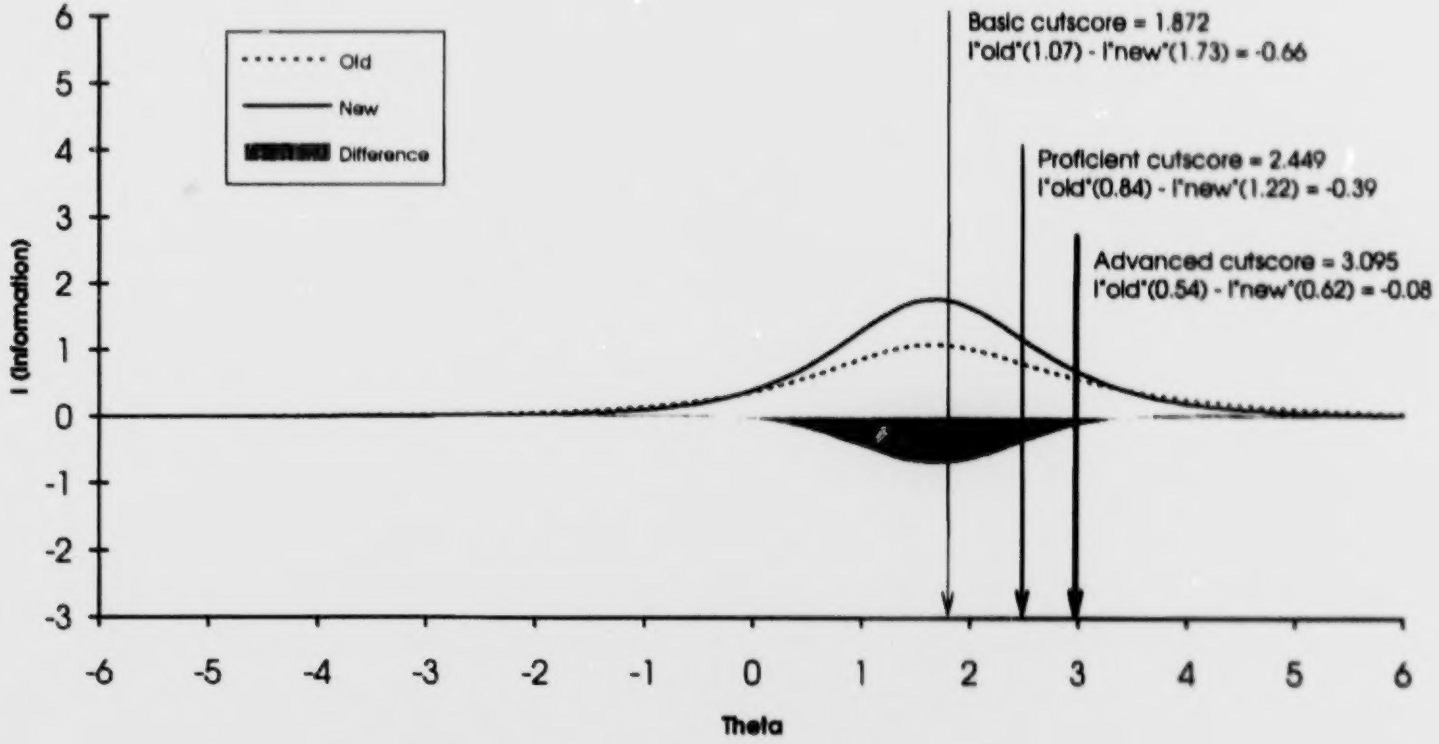


Table K-4

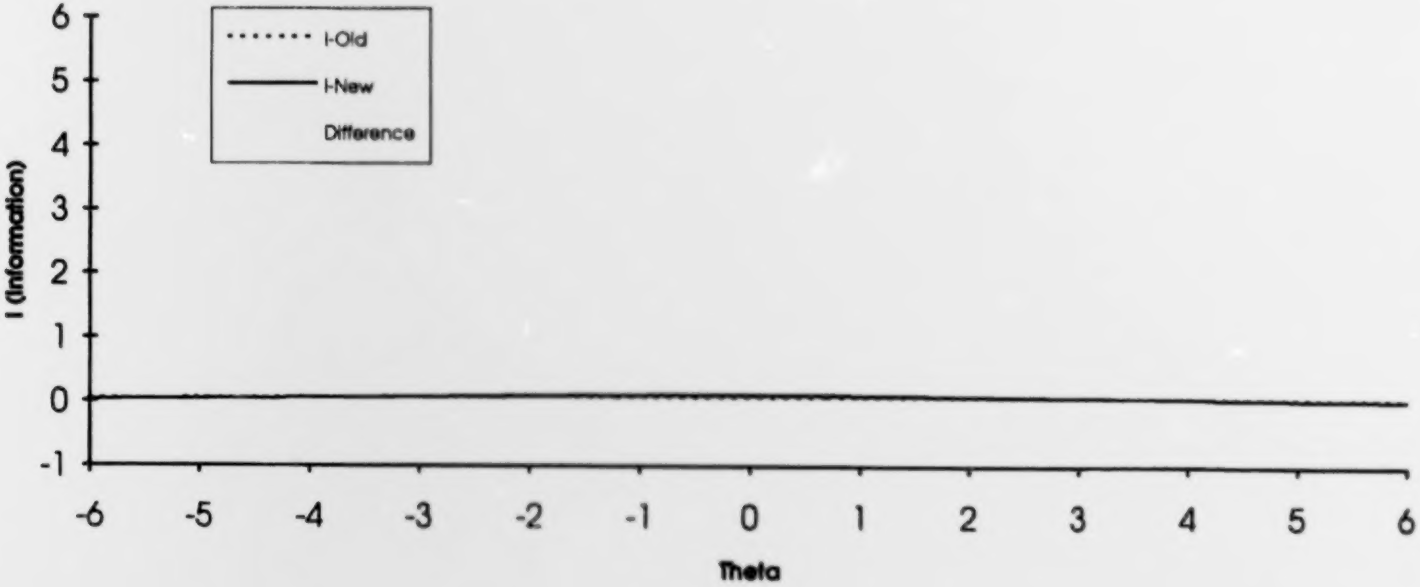
Comparison Between the "Old" and "New" Information Weights for 1992 Reading Achievement Levels Based on Corrected Item Parameters

Grade	Rating Group	Scale	Basic			Proficient			Advanced		
			Old	New	Diff	Old	New	Diff	Old	New	Diff
4	A	1	4.79	3.40	1.39	3.44	2.07	1.37	3.25	1.72	1.53
		2	1.15	1.02	0.13	0.87	0.60	0.27	0.61	0.35	0.26
	B	1	4.34	4.17	0.17	3.49	3.02	0.47	1.85	1.22	0.53
		2	1.84	1.48	0.36	2.88	2.75	0.13	2.78	2.62	0.17
8	A	1	4.34	3.73	0.61	4.15	3.37	0.78	3.30	1.94	1.36
		2	3.12	2.76	0.37	3.10	2.65	0.45	2.17	1.63	0.55
		3	0.77	0.81	-0.04	0.85	1.08	-0.24	0.68	0.72	-0.04
	B	1	1.68	1.55	0.13	1.15	0.70	0.46	0.81	0.42	0.40
		2	3.51	3.90	-0.40	3.58	3.92	-0.35	3.27	3.20	0.08
		3	0.66	0.77	-0.11	0.66	0.77	-0.11	0.59	0.82	-0.23
12	A	1	1.85	1.38	0.47	1.65	1.22	0.43	1.99	1.13	0.86
		2	4.94	5.75	-0.81	4.17	4.30	-0.12	2.18	1.75	0.43
		3	0.81	1.05	-0.24	0.53	0.57	-0.04	0.47	0.48	-0.01
	B*	2	6.31	5.19	1.13	3.82	2.35	1.47	1.51	0.84	0.67
		3	1.07	1.73	-0.66	0.84	1.22	-0.39	0.54	0.62	-0.08

****Note***: *No polytomous items in Scale 1 were rated by panelists in Group B.*

Figure K-12

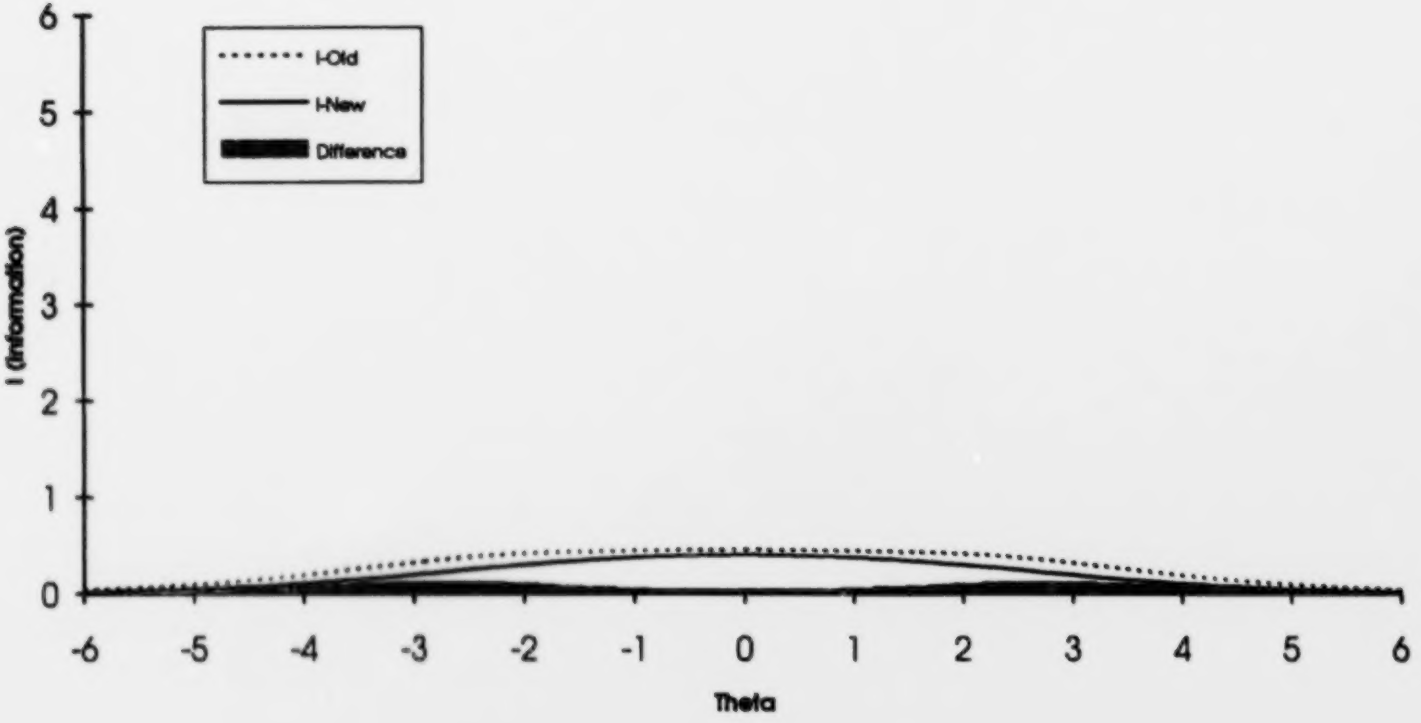
Comparison Between the "Old" and "New" Information Functions Using Hypothetical Item Parameters a=0.2, b=0, d0=0, d1=2, and d3=-2, in Case of Four Response Categories



993

994

Figure K-13
Comparison Between the "Old" and "New" Information Functions Using Hypothetical Item Parameters a=0.5, b=0, d0=0, d1=2, and d3=-2, in Case of Four Response Categories

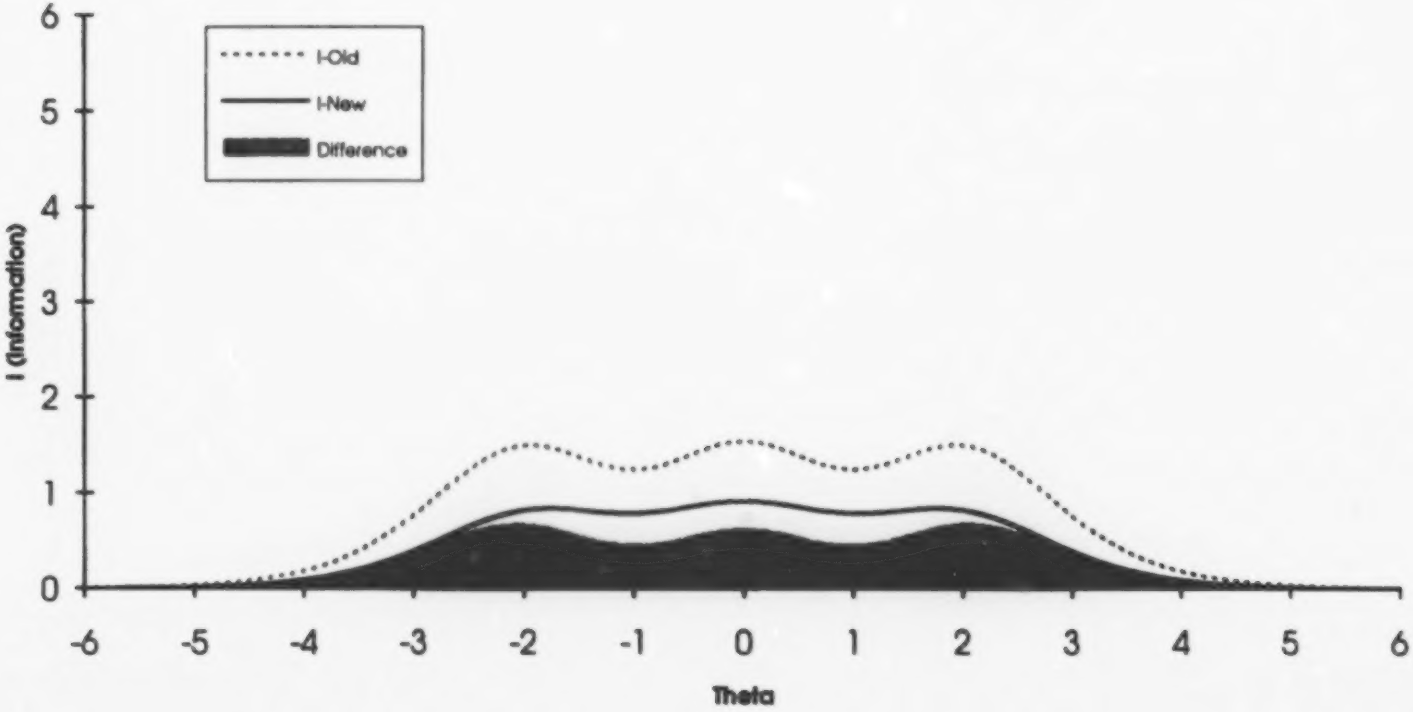


995

996

Figure K-14

Comparison Between the "Old" and "New" Information Functions
Using Hypothetical Item Parameters a=1, b=0, d0=0, d1=2, d2=0, and d3=-2,
in Case of Four Response Categories



Analysis of the Effect of Location Parameters. Figures K-15, K-16, and K-17 show the correct and incorrect information weighting functions for varying location parameters. The location parameters only shift the distribution of information, and that is the case for both the correct and incorrect information weighting functions. The amount of difference between the two is unchanged; only the locations change.

Analysis of the Effect of the Threshold Parameters. If the threshold parameters are close, in terms of the locations of ICCs, the correct information function will have a high peak. In the area of maximum information, i.e., around the peak of the distribution, the incorrect information function underestimates information.

As can be seen in Figures K-15 and K-15a, when the threshold parameters are relatively far apart, the information is relatively low and the distribution is multimodal. When the threshold parameters are closer (see Figures K-18 and K-18a) maximum information is higher and the distribution tends to be more unimodal. As the threshold parameters move even closer, the difference between the correct and incorrect information functions decreases. (See Table K-5.) As the threshold parameters become even closer (see Figures K-19, K-19a, K-20, and K-20a), the rate at which the incorrect weighting function increments at the peak of the distribution is slower than that for the correct function. This results in a negative difference between the two functions in the area of maximum information. (See Figures K-20 and K-20a.)

Figure K-10 is an example of how the location of the threshold parameters can impact the weights of the polytomous cutpoints. The Basic cutpoint falls where the incorrect function weights are less than the correct function weights, but the Advanced cutpoint falls where the incorrect function weights are greater than the correct function weights. Thus, the impact of the incorrect information function is seen to be inconsistent and to depend upon the location of the cutpoint.

Analysis of the Error: Conclusions

Three general conclusions can be drawn from our analyses.

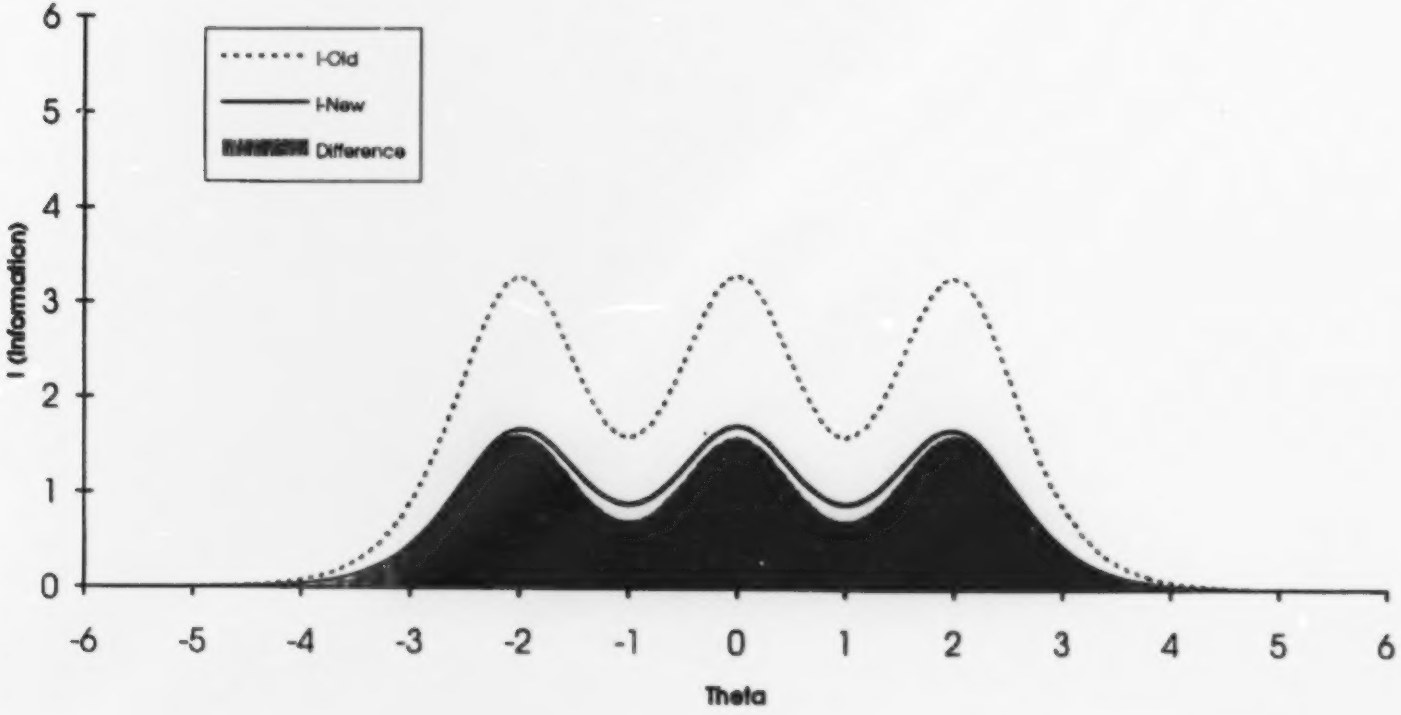
1. Generally, the incorrect information weighting function results in a higher information weight. This is not a consistent pattern, however, because the incorrect function can result in lower information weights than the correct function.

2. The difference between the correct and incorrect weights increases as item discrimination increases.

3. The impact of the incorrect information weighting function on the cutscores is not consistent. The impact depends upon the location of the cutscore and the relative weight of the dichotomous items.

Figure K-15

Comparison Between the "Old" and "New" Information Functions
Using Hypothetical Item Parameters a=1.5, b=0, d0=0, d1=2, d2=0, and d3=-2,
in Case of Four Response Categories



1000

1001

Figure K-15a

ICC of a Polytomous Item with a=1.5, b=0, d0=0, d1=2, d3=-2

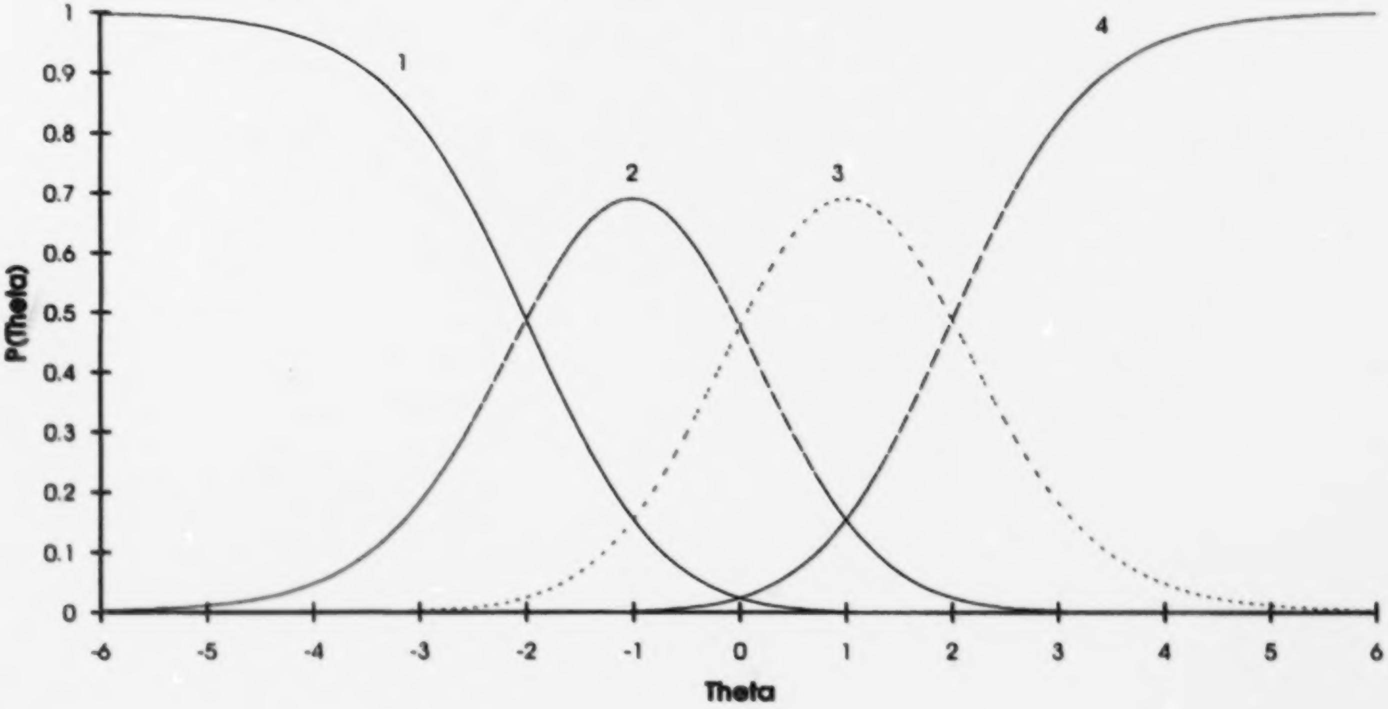
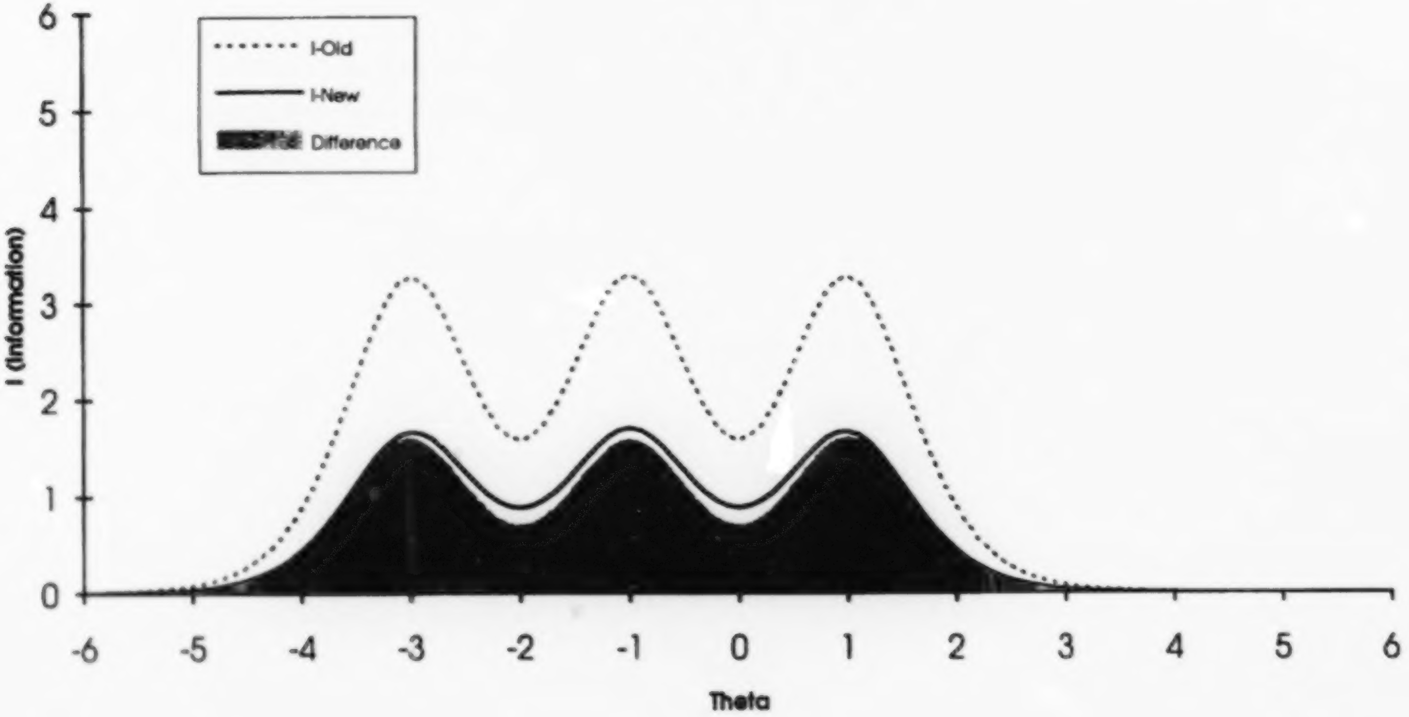


Figure K-16

Comparison Between the "Old" and "New" Information Functions Using Hypothetical Item Parameters a=1.5, b=-1, d0=0, d1=2, d2=0, and d3=-2 in Case of Four Response Categories



1004

100

Figure K-17

Comparison Between the "Old" and "New" Information Functions
Using Hypothetical Item Parameters a=1.5, b=1, d0=0, d1=2, d2=0, and d3=-2
in Case of Four Response Categories

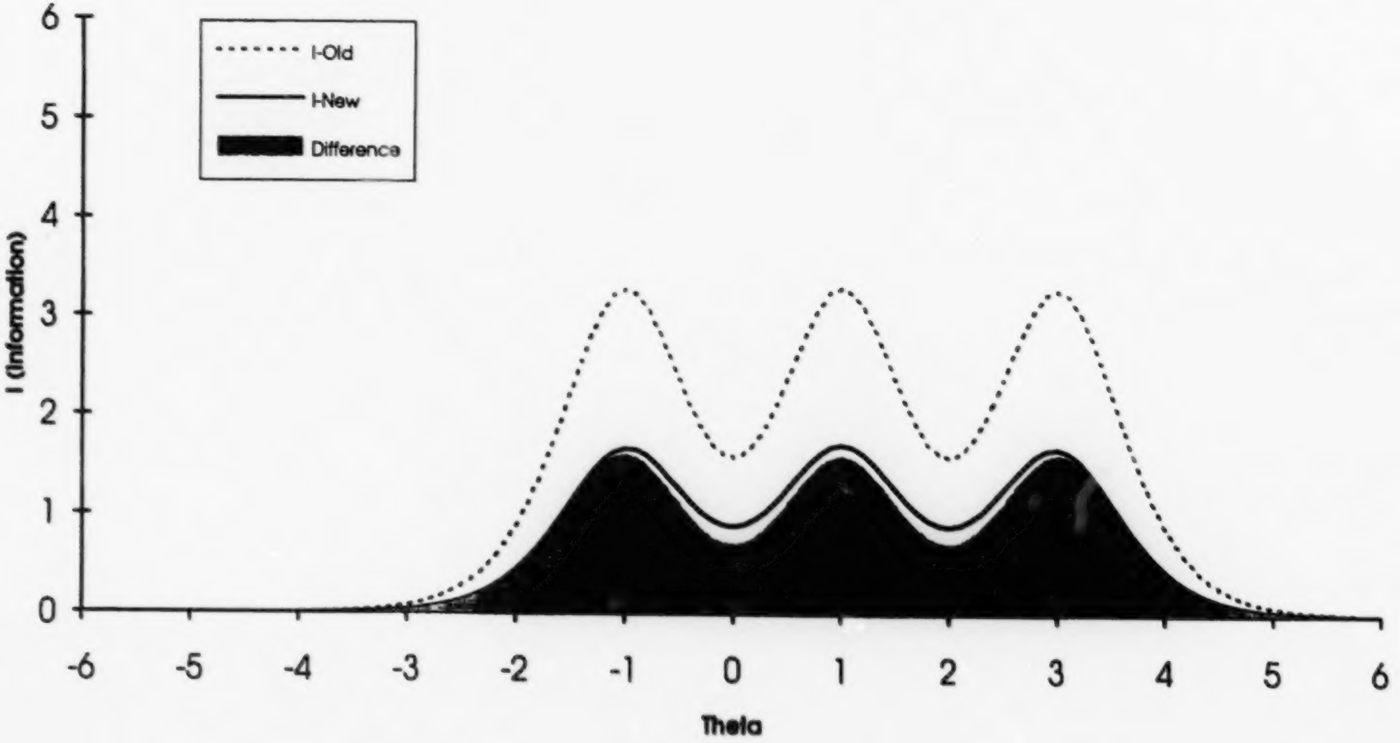


Figure K-18

Comparison Between the "Old" and "New" Information Functions
Using Hypothetical Item Parameters a=1.5, b=0, d0=0, d1=2, d2=0, and d3=-1
in Case of Four Response Categories

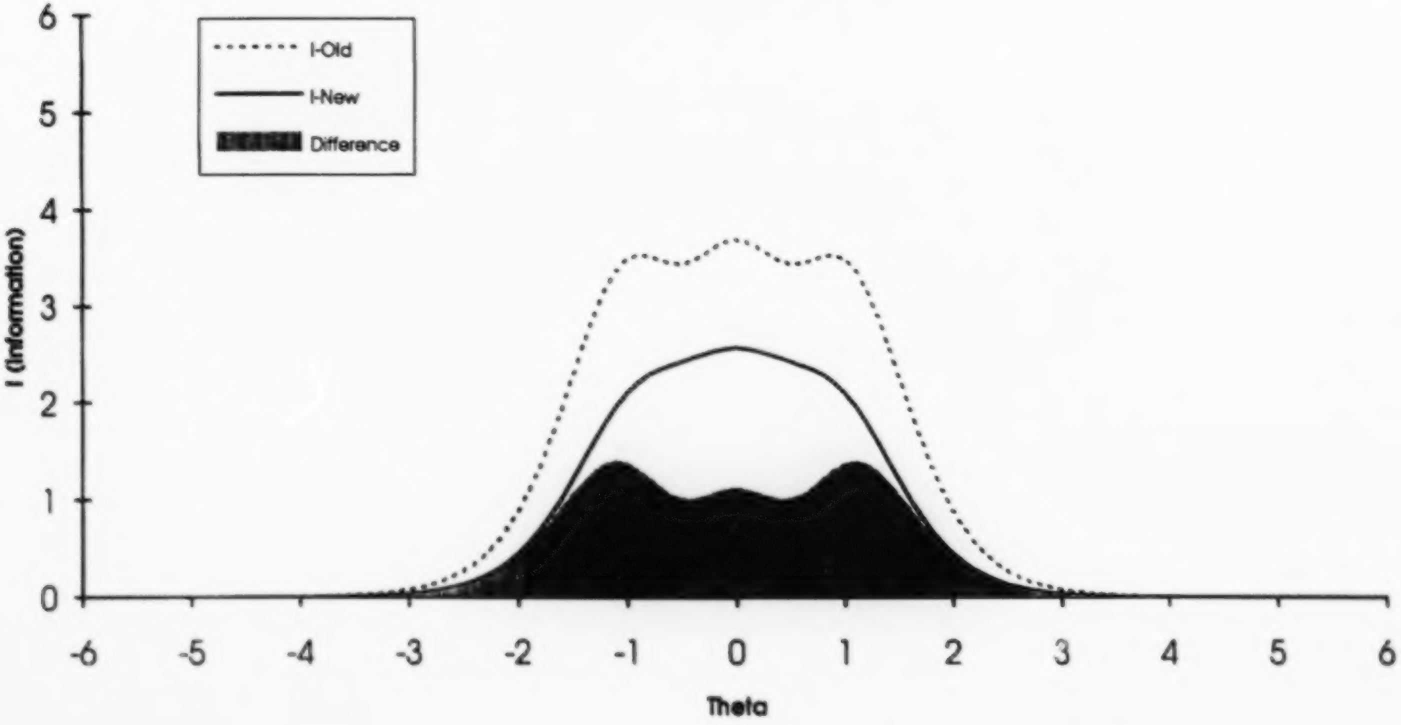
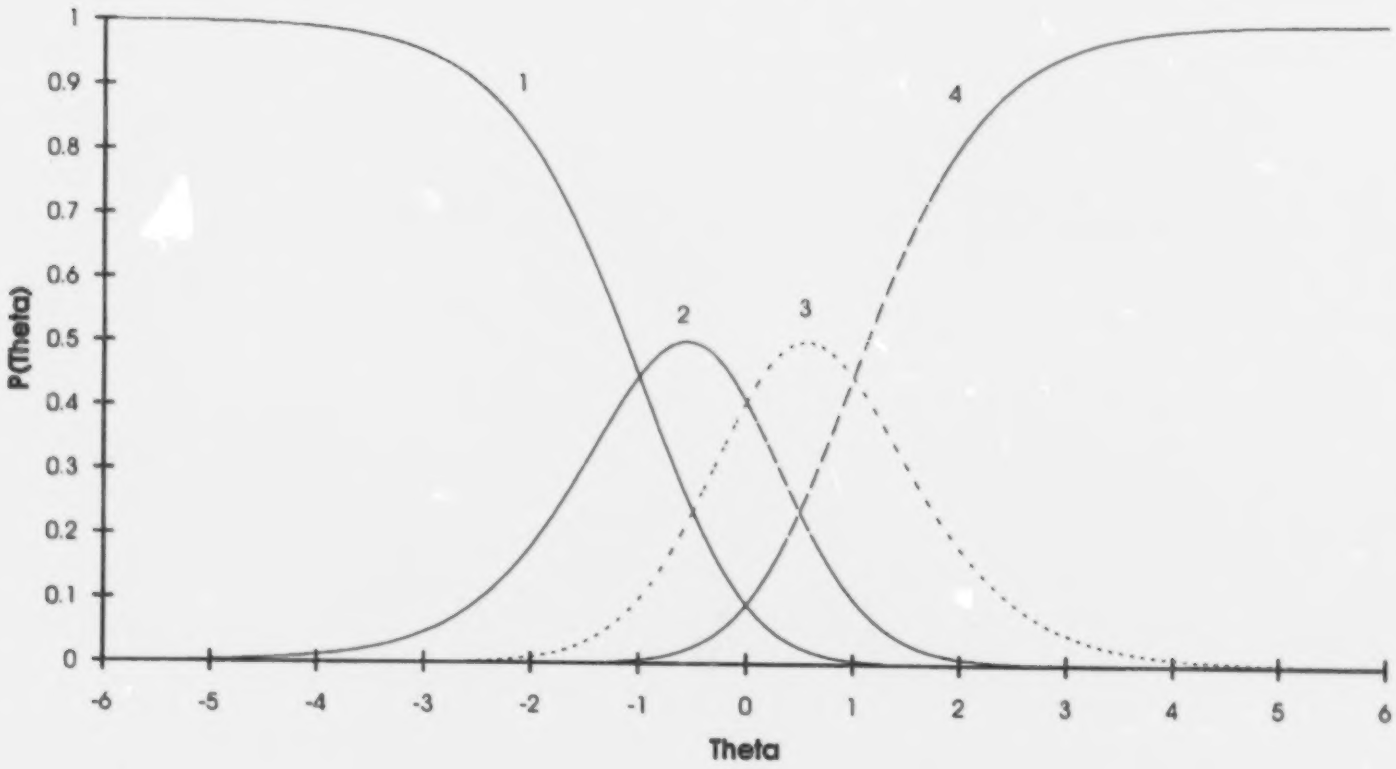


Figure K-18a

ICC of a Polytomous Item with a=1.5, b=0, d0=0, d1=1, d2=0, and d3=-1



1010

1011

Table K-5

Item Parameters Used for Figures K-12 through K-20 and the Sum of the Difference Between the "Old" and the "New" Information Weighting Functions

Figure	a	b	d_0	d_1	d_2	d_3	Sum
12	0.2	0	0	2	0	-2	1.37
13	0.5	0	0	2	0	-2	9.67
14	1.0	0	0	2	0	-2	38.52
15	1.5	0	0	2	0	-2	70.47
16	1.5	-1	0	2	0	-2	70.46
17	1.5	1	0	2	0	-2	70.40
18	1.5	0	0	1	0	-1	46.30
19	1.5	0	0	.52	0	-.52	22.40
20	1.5	0	0	.1	0	-.1	21.46

Figure K-19

Comparison Between the "Old" and "New" Information Functions
Using Hypothetical Item Parameters a=1.5, b=0, d0=0, d1=0.52, d2=0, and d3=-0.52
in Case of Four Response Categories

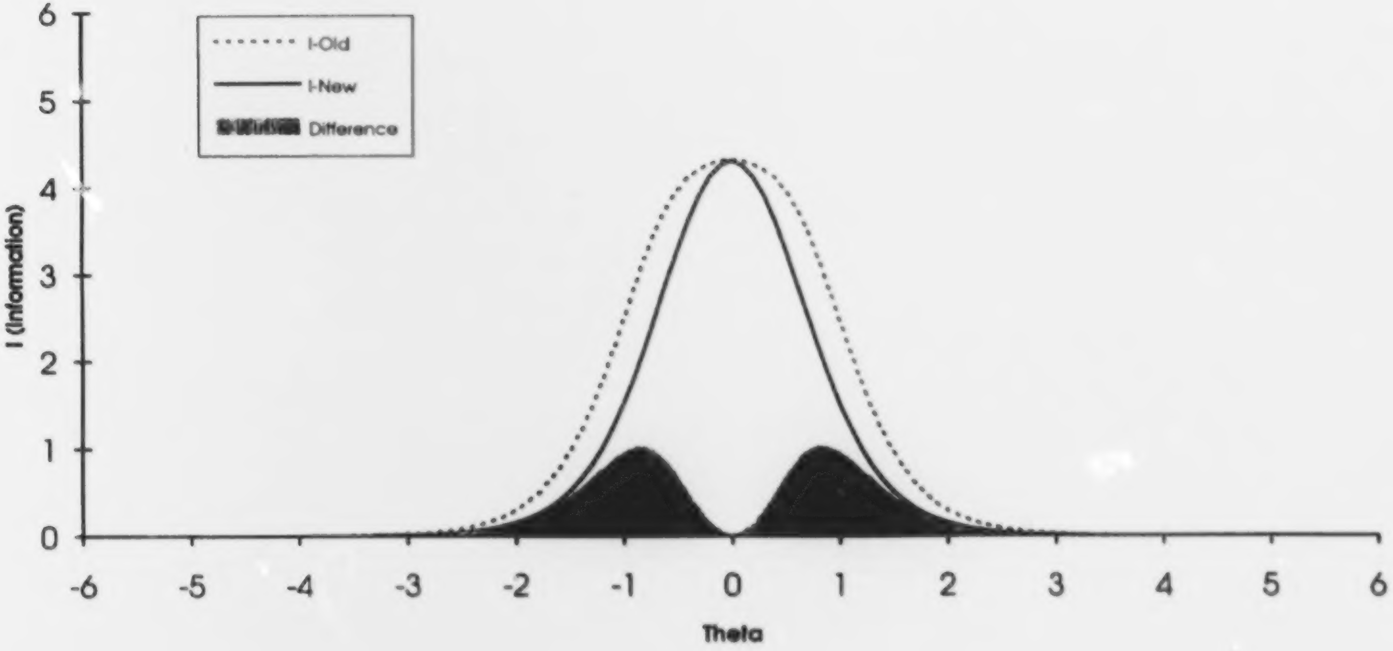


Figure K-19a

Item Characteristic Curves of a=1.5, b=0, d0=0, d1=0.52, d2=0, and d3=-0.52

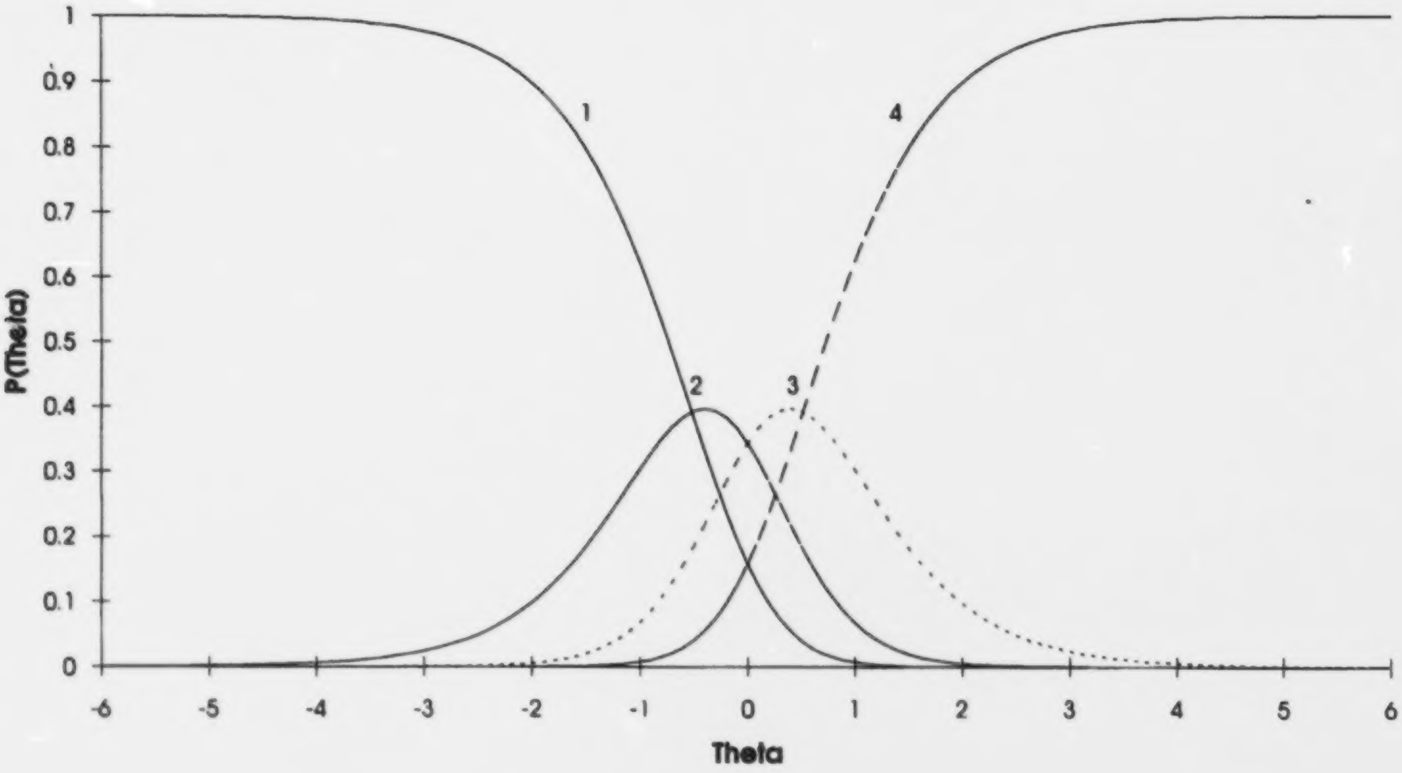


Figure K-20

Comparison Between the "Old" and "New" Information Functions
Using Hypothetical Item Parameters a=1.5, b=0, d0=0, d1=0.1, d2=0, and d3=-0.1
in Case of Four Response Categories

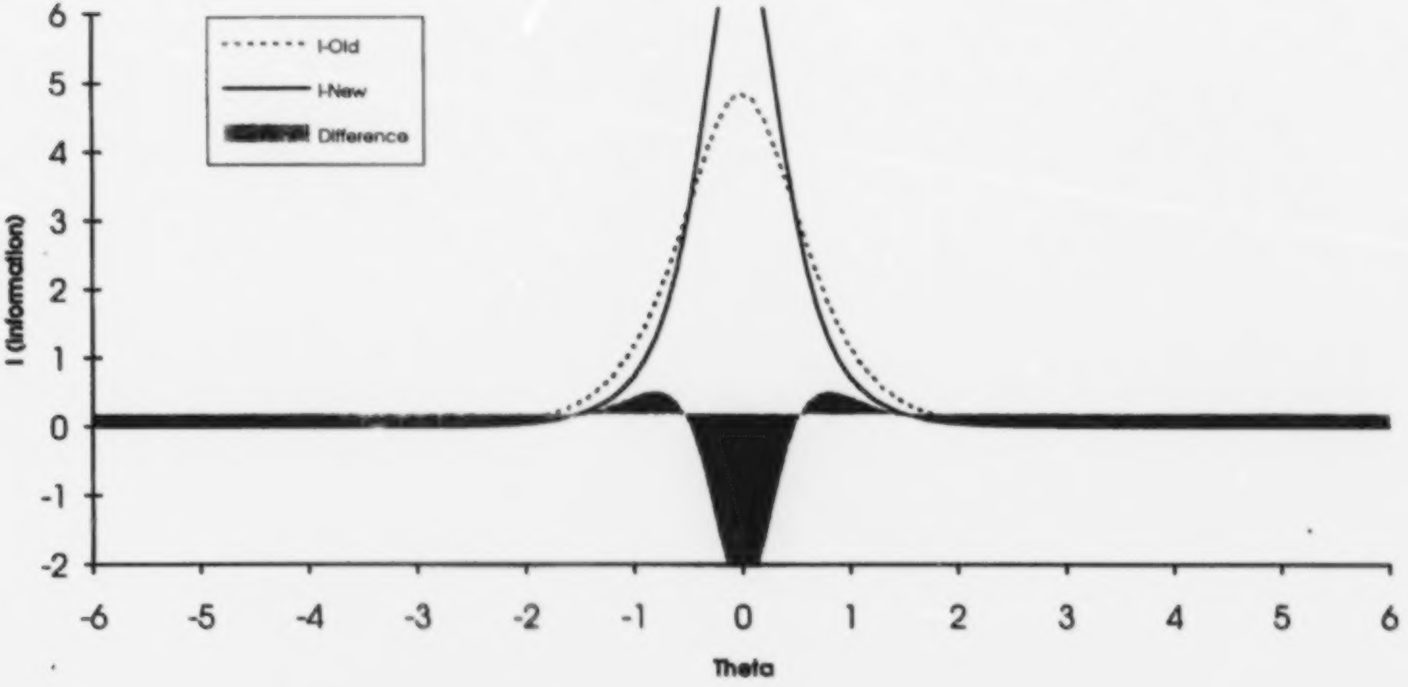
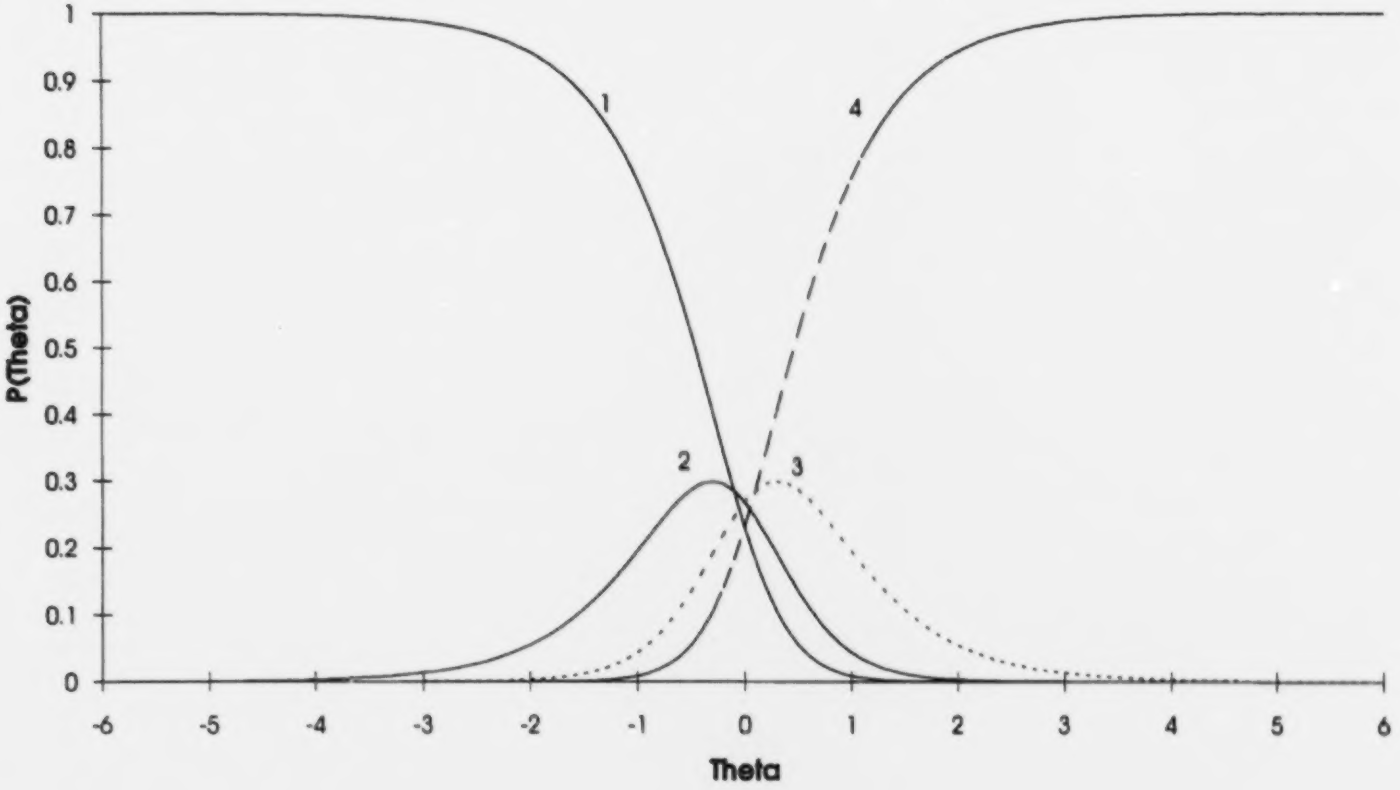


Figure K-20a

Item Characteristic Curves of a=1.5, b=0, d0=0, d1=0.1, d2=0, and d3=-0.1



REFERENCES CITED IN TEXT

REFERENCES CITED IN TEXT

Akaike, H. (1987). Factor analysis and AIC. *Psychometrika, 52*, 317-332.

Allen, N. L. (1992). Data analysis for the science assessment. In E. G. Johnson & N. L. Allen, *The NAEP 1990 technical report* (No. 21-TR-20). Princeton, NJ: Educational Testing Service, National Assessment of Educational Progress.

Allen, N. L., & Donoghue, J. R. (1994, April). *Differential item functioning based on complex samples of dichotomous and polytomous items*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Allen, N. L., & Donoghue, J. R. (in press). Applying the Mantel-Haenszel procedure to complex samples of items. *Journal of Educational Measurement.*

Allen, N. L., & Isham, S. P. (1994). Data analysis for the science long-term trend assessment. In E. G. Johnson & J. E. Carlson, *The NAEP 1992 technical report.* Washington, DC: National Center for Education Statistics.

American College Testing. (1993). *Setting achievement levels on the 1992 National Assessment of Educational Progress in mathematics, reading, and writing: A technical report on reliability and validity.* Iowa City, IA: Author.

American College Testing. (1995a). *Research studies on the achievement levels set for the 1994 NAEP in geography and U.S. history.* Iowa City, IA: Author.

American College Testing. (1995b). *Results of the 1994 geography NAEP achievement levels-setting pilot study, July 1994.* Iowa City, IA: Author.

American College Testing. (1995c). *Results of the 1994 U.S. history NAEP achievement levels-setting pilot study, August 1994.* Iowa City, IA: Author.

Andersen, E. B. (1980). Comparing latent distributions. *Psychometrika, 45*, 121-134.

Angoff, W. H. (1971). Scales, norms, and equivalent scores. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed., pp. 508 - 600). Washington, DC: American Council on Education.

Beaton, A. E. (1987). *Implementing the new design: The NAEP 1983-84 technical report.* (No. 15-TR-20) Princeton, NJ: Educational Testing Service, National Assessment of Educational Progress.

Beaton, A. E., & Johnson, E. G. (1990). The average response method of scaling. *Journal of Educational Statistics, 15*, 9-38.

Beaton, A. E., & Johnson, E. G. (1992). Overview of the scaling methodology used in the National Assessment. *Journal of Educational Measurement, 26(2)*, 163-175.

Beaton, A. E., & Zwick, R. (1990). *The effect of changes in the National Assessment: Disentangling the NAEP 1985-86 reading anomaly.* (No. 17-TR-21) Princeton, NJ: Educational Testing Service, National Assessment of Educational Progress.

Beatty, A. S., Reese, C. M., Persky, H. R., & Carr, P. (1996). *NAEP 1994 U.S. history report card: Findings from the National Assessment of Educational Progress.* Washington, DC: National Center for Education Statistics.

Bock, R. D. (1972). Estimating item parameters and latent ability when responses are scored in two or more latent categories. *Psychometrika, 37,* 29-51.

Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: application of an EM algorithm. *Psychometrika, 46,* 443-459.

Bock, R. D., Gibbons, R., & Muraki, E. (1988). Full-information item factor analysis. *Applied Psychological Measurement, 12,* 261-280.

Bourque, M. L., & Garrison, H. H. (1991). *The levels of mathematics achievement. Vol. I, national and state summaries.* Washington, DC: National Assessment Governing Board.

Campbell, J. R., Donahue, P. L., Reese, C. M., & Phillips, G. W. (1996). *NAEP 1994 reading report card for the nation and the states: Findings from the National Assessment of Educational Progress and Trial State Assessments.* Washington, DC: National Center for Education Statistics.

Campbell, J. R., Reese, C. M., O'Sullivan, C., Dossey, J. A., & Donahue, P. L. (1996). *NAEP 1994 trends in academic progress: Achievement of U.S. students in Science, 1969 to 1994; mathematics, 1973 to 1994; reading, 1971 to 1994; and writing, 1984 to 1994.* Washington, DC: National Center for Education Statistics.

Carlson, J. E. (1993). *Dimensionality of NAEP instruments that incorporate polytomously-scored items.* Paper presented at the annual meeting of the American Educational Research Association, Atlanta, GA.

Carlson, J. E., & Jirele, T. (1992, April). *Dimensionality of 1990 NAEP mathematics data.* Paper presented at the meeting of the American Educational Research Association, San Francisco, CA.

Cizek, G. (1993). *Reactions to National Academy report, "Setting performance standards for student achievement".* Washington, DC: National Assessment Governing Board.

Cohen, J. (1968). Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin, 70*(4), 213-220.

Curry, L. (1987, April). *Group decision process in setting cut-off scores.* Paper presented at the annual meeting of the American Educational Research Association, Washington, DC.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society, Series B, 39,* 1-38.

Donoghue, J. R. (1995, April). *Assessing some of the measurement properties of longer reading blocks: An application of the bootstrap to structural equation models.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Donoghue, J. R., Isham, S. P., Bowker, D. W., & Freund, D. S. (1994). Data analysis for the reading assessment. In E. G. Johnson & J. E. Carlson, *The NAEP 1992 technical report.* Washington, DC: National Center for Education Statistics.

Donoghue, J. R., & Mazzeo, J. (1993). *Assessing the properties of longer blocks in the 1992 reading assessment.* Paper presented at the symposium "Analysis of Innovative Data Structures in the National Assessment of Educational Progress" at the annual meeting of the American Educational Research Association, Atlanta, GA.

Educational Testing Service (1987). *ETS standards for quality and fairness.* Princeton, NJ: Author.

Engelen, R. J. H. (1987). *Semiparametric estimation in the Rasch model.* Research Report 87-1. Twente, the Netherlands: Department of Education, University of Twente.

Fitzpatrick, A. R. (1989). Social influences in standard-setting: The effects of social interaction on group judgments. *Review of Educational Research, 59,* 315-328.

Fraser, C. (1988). *NOHARM A computer program for fitting both unidimensional and multidimensional normal ogive models of latent trait theory.* Armidale, Australia: The University of New England, Center for Behavioral Studies.

Friedman, C. B., & Ho, K. T. (1990, April). *Interjudge consensus and intrajudge consistency: Is it possible to have both on standard setting?* Paper presented at the annual meeting of the National Council for Measurement in Education, Boston, MA.

Hambleton, R. K., & Bourque, M. L. (1991). *The levels of mathematics achievement. Vol. II, technical report.* Washington, DC: National Assessment Governing Board.

Hoijtink, H. (1991). *Estimating the parameters of linear models with a latent dependent variable by nonparametric maximum likelihood.* Research Bulletin HB-91-1040-EX. Groningen, The Netherlands: Psychological Institute, University of Groningen.

Holland, P. W., & Thayer, D. T. (1988). Differential item performance and the Mantel-Haenszel procedure. In H. Wainer & H. I. Braun (Eds.), *Test validity.* Hillsdale, NJ: Erlbaum.

Huynh, H. (1994, October). *Some technical aspects of standard setting.* Paper presented at the Joint Conference on Standard Setting for Large-Scale Assessments, Washington, DC.

Johnson, E. G. (1989). Considerations and techniques for the analysis of NAEP data. *Journal of Educational Statistics, 14,* 303-334.

Johnson, E. G. (1990). Data analysis for the writing assessment. In E. G Johnson & R. Zwick (Eds.), *Focusing the new design: the NAEP 1988 technical report* (pp. 267-296). Washington, DC: National Center for Education Statistics, Office of Educational Research and Improvement, U.S. Department of Education.

Johnson, E. G. (1993). Special studies on constructed response scoring. In Educational Testing Service, *The results of the NAEP 1993 field test for the 1994 National Assessment of Educational Progress* (pp. 23-58). Princeton, NJ: Educational Testing Service, National Assessment of Educational Progress.

Johnson, E. G., & Allen, N. L. (1992). *The NAEP 1990 technical report* (No. 21-TR-20). Princeton, NJ: Educational Testing Service, National Assessment of Educational Progress.

Johnson, E. G., & Carlson, J. E. (1994). *The NAEP 1992 technical report.* Washington, DC: National Center for Education Statistics.

Johnson, E. G., & King, B. F. (1987). Generalized variance functions for a complex sample survey. *Journal of Official Statistics, 3*, 235-250.

Johnson, E. G., & Freund, D. S. (1994, April). *Scoring of performance items using image processing: Comparability study with paper-and-pencil approach.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Johnson, E. G., & Rust, K. F. (1992). Population inferences and variance estimation for NAEP data. *Journal of Educational Statistics, 17*, 175-190.

Johnson, E. G., & Zwick, R. (1990). *Focusing the new design: The NAEP 1988 technical report.* (No. 19-TR-20) Princeton, NJ: Educational Testing Service, National Assessment of Educational Progress.

Johnson, J. R. (1987). Instrument and item information. In A. E. Beaton, *Implementing the new design: The NAEP 1983-84 technical report* (pp. 119-134). (No. 15-TR-20) Princeton, NJ: Educational Testing Service, National Assessment of Educational Progress.

Kane, M. (1993). *Comments on the NAE evaluation of the NAGB achievement levels.* Washington, DC: National Assessment Governing Board.

Kaplan, B. A., Beaton, A. E., Johnson, E. G., & Johnson, J. R. (1988). *National Assessment of Educational Progress: 1986 Bridge Studies.* Princeton. NJ: Educational Testing Service, National Assessment of Educational Progress.

Kaplan, D. (1995). The impact of BIB spiraling-induced missing data patterns on goodness-of-fit tests in factor analysis. *Journal of Educational and Behavioral Statistics, 20*, 69-82.

Kish, L. (1965). *Survey sampling*. New York, NY: John Wiley & Sons.

Kish, L., & Frankel, M. R. (1974). Inference from complex samples. *Journal of the Royal Statistical Society, Series B, 36*, 1-22.

Krewski, D., & Rao, J. N. K. (1981). Inference from stratified samples: Properties of linearization, jackknife and balanced repeated replication. *Annals of Statistics, 9*, 1010-1019.

Laird, N. M. (1978). Nonparametric maximum likelihood estimation of a mixing distribution. *Journal of the American Statistical Association, 73*, 805-811.

Langer, J. A. (1989). *The process of understanding literature*. (Technical report No. 2.1.) Albany, NY: State University of New York, Center for the Learning and Teaching of Literature.

Langer, J. A. (1990). The process of understanding: Reading for literary and informative purposes. *Research in the Teaching of English, 24*, 229-257.

Lindsey, B., Clogg, C. C., & Grego, J. (1991). Semiparametric estimation in the Rasch model and related exponential response models, including a simple latent class model for item analysis. *Journal of the American Statistical Association, 86*, 96-107.

Little, R. J. A., & Rubin, D. B. (1983). On jointly estimating parameters and missing data. *American Statistician, 37*, 218-220.

Little, R. J. A., & Rubin, D. B. (1987). *Statistical analysis with missing data*. New York, NY: John Wiley & Sons.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Mantel, N. (1963). Chi-square tests with one degree of freedom: Extensions of the Mantel-Haenszel procedure. *Journal of the American Statistical Association, 58*, 690-700.

Mazzeo, J., Allen, N. L., & Kline, D. L. (1995). *Technical report of the NAEP 1994 Trial State Assessment program in reading*. Washington, DC: National Center for Education Statistics.

Mazzeo, J., Johnson, E. G., Bowker, D., & Fong, Y. F. (1992). *The use of collateral information in proficiency estimation for the Trial State Assessment*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

McDonald, R. P. (1962). A general approach to nonlinear factor analysis. *Psychometrika, 27*, 397-415.

McDonald, R. P. (1967). Nonlinear factor analysis. *Psychometric Monographs*, No. 15.

McDonald, R. P. (1981). The dimensionality of tests and items. *British Journal of Mathematical and Statistical Psychology, 34*, 100-117.

McDonald, R. P. (1982a). Linear versus nonlinear models in item response theory. *Applied Psychological Measurement, 6*, 379-396.

McDonald, R. P. (1982b). Unidimensional and multidimensional models for item response theory. In D. J. Weiss (Ed.), *Proceedings of the 1982 Computerized Adaptive Testing Conference* (pp. 127-148). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program.

McDonald, R. P. (1985). *Factor analysis and related methods.* Hillsdale, NJ: Lawrence Erlbaum.

Messick, S. J., Beaton, A. E., & Lord, F. M. (1983). *NAEP reconsidered: A new design for a new era.* (NAEP Report 83-1) Princeton, NJ: Educational Testing Service.

Miller, K. E., Nelson, J. E., & Naifeh, M. (1995). *Cross-state data compendium for the NAEP 1994 grade 4 reading assessment.* Washington, DC: National Center for Education Statistics.

Mislevy, R. J. (1985). Estimation of latent group effects. *Journal of the American Statistical Association, 80*, 993-997.

Mislevy, R. J. (1990). Scaling procedures. In E. G. Johnson and R. Zwick, *Focusing the new design: The NAEP 1988 technical report* (No. 19-TR-20). Princeton, NJ: Educational Testing Service, National Assessment of Educational Progress.

Mislevy, R. J. (1991). Randomization-based inference about latent variables from complex samples. *Psychometrika, 56*, 177-196.

Mislevy, R. J., Beaton, A. E., Kaplan, B., & Sheehan, K. M. (1992). Estimating population characteristics from sparse matrix samples of item responses. *Journal of Educational Measurement, 29*(2), 133-161.

Mislevy, R. J., & Bock, R. D. (1982). *BILOG: Item analysis and test scoring with binary logistic models* [Computer program]. Mooresville, IN: Scientific Software.

Mislevy, R. J., Johnson, E. G., & Muraki, E. (1992). Scaling procedures in NAEP. *Journal of Educational Statistics, 17*(2), 131-154.

Mislevy, R. J., & Sheehan, K. M. (1987). Marginal estimation procedures. In A. E. Beaton, *Implementing the new design: The NAEP 1983-84 technical report* (No. 15-TR-20). Princeton, NJ: Educational Testing Service, National Assessment of Educational Progress.

Mislevy, R. J., & Wu, P-K. (1988). *Inferring examinee ability when some item responses are missing* (ETS Research Report RR-88-48-ONR). Princeton, NJ: Educational Testing Service.

Mullis, I. V. S., Campbell, J. R., & Farstrup, A. E. (1993). *NAEP 1992 reading report card for the nation and the states. Data from the national and Trial State assessments.* Washington, DC: National Center for Education Statistics.

Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement, 16*(2), 159-176.

Muraki, E., & Bock, R. D. (1991). *PARSCALE: Parameter scaling of rating data*. Chicago, IL: Scientific Software, Inc.

Muraki, E. & Grima, A. M. (1993, May). *Extension of TBLT procedure to the generalized partial credit model*. Paper presented at the NAEP Design and Analysis Committee Meeting, Washington, DC.

Muthén, B. (1991). *Issues in using NAEP mathematics items to study achievement dimensionality, within-grade differences, and across-grade growth*. Report presented to the Design and Analysis Committee for NAEP.

National Academy of Education. (1992). *Assessing student achievement in the states. The first report of the National Academy of Education Panel on the evaluation of the NAEP Trial State Assessment: 1990 Trial State Assessment*. Stanford, CA: Author.

National Academy of Education. (1993a). *Setting performance standards for student achievement*. Stanford, CA: Author.

National Academy of Education. (1993b). *The trial state assessment: prospects and realities*. Stanford, CA: Author.

National Assessment of Educational Progress. (1996). *The NAEP guide: A description of the content and methods of the 1994 and 1996 assessments*. Washington, DC: National Center for Education Statistics.

National Assessment Governing Board (1989). *Setting achievement goals on the National Assessment of Educational Progress: A draft policy statement*. Washington, DC: Author.

National Assessment Governing Board (1990). *Setting appropriate achievement levels for the National Assessment of Educational Progress: Policy framework and technical procedures*. Washington, DC: Author.

National Assessment Governing Board (1994). *Geography framework for the 1994 National Assessment of Educational Progress*. Washington, DC: Author.

National Assessment Governing Board. (1994). *Reading framework for the 1992 and 1994 National Assessment of Educational Progress*. Washington, DC: Author.

National Assessment Governing Board. (1994). *U.S. history framework for the 1994 National Assessment of Educational Progress*. Washington, DC: Author.

National Computer Systems. (1995). *Report of processing and professional activities*. Iowa City, IA: Author.

O'Reilly, P. O., Zelenak, C. A., Rogers, A. M., & Kline, D. L. (1996a). *NAEP 1994 Trial State Assessment program in reading secondary-use data files user guide*. Princeton, NJ: Educational Testing Service, National Assessment of Educational Progress.

O'Reilly, P. O., Zelenak, C. A., Rogers, A. M., & Kline, D. L. (1996b). *NAEP 1994 secondary-use data files user guide*. Princeton, NJ: Educational Testing Service, National Assessment of Educational Progress.

Persky, H. R., Reese, C. M., O'Sullivan, C. Y., Lazer, S., Moore, J., & Shakrani, S. (1996). *NAEP 1994 geography report card: Findings from the National Assessment of Educational Progress*. Washington, DC: National Center for Education Statistics.

Petersen, N. (1988). *DIF procedures for use in statistical analysis*. Internal memorandum.

Rock, D. A. (1991, November). *Subscale dimensionality*. Paper presented at the meeting of the Design and Analysis Committee of the National Assessment of Educational Progress, Washington, DC.

Rogers, A. M., & Nelson, J. E. (1990). *NAEP-BILOG: Enhanced version of Mislevy and Bock's software for item analysis and test scoring with binary logistic models* [Computer program]. Princeton, NJ: Educational Testing Service.

Rosenbaum, P. R. (1984). Testing the conditional independence and monotonicity assumptions of item response theory. *Psychometrika, 49*, 425-436.

Rosenbaum, P. R. (1985). Comparing distributions of item responses for two groups. *British Journal of Mathematical and Statistical Psychology, 38*, 206-215.

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York, NY: John Wiley & Sons.

Rubin, D. B. (1991). EM and beyond. *Psychometrika, 56*, 241-254.

Rust, K. (1985). Variance estimation for complex estimators in sample surveys. *Journal of Official Statistics, 1*(4), 381-397.

Rust, K. F., Bethel, J., Burke, J., & Hansen, M. H. (1990). *The 1988 National Assessment of Educational Progress—Sampling and weighting procedures, final report*. Rockville, MD: Westat, Inc.

Rust, K. F., Burke, J., & Fahimi, M. (1992). *1990 National Assessment of Educational Progress sampling and weighting procedures, part 2: National assessment, final report*. Rockville, MD: Westat, Inc.

Sheehan, K. M. (1985). *MGROUP: Estimation of group effects in multivariate models* [Computer program]. Princeton, NJ: Educational Testing Service.

Smith, R. L., & Smith, J. K. (1988). Differential use of item information by judges using Angoff and Nedelsky procedures. *Journal of Educational Measurement, 25*, 259-274.

Somes, G. W. (1986). The generalized Mantel-Haenszel statistic. *The American Statistician, 40*, 106-108.

Stocking, M. L., & Lord, F. M. (1983). Developing a common metric in item response theory. *Applied Psychological Measurement, 7*, 201-210.

Tanner, M., & Wong, W. (1987). The calculation of posterior distributions by data augmentation (with discussion). *Journal of the American Statistical Association, 82*, 528-550.

Thomas, N. (1993). Asymptotic corrections for multivariate posterior moments with factored likelihood functions. *Journal of Computational and Graphical Statistics, 2*, 309-322.

Wallace, L., & Rust, K. F. (1996). *The 1994 NAEP sampling and weighting report*. Rockville, MD: Westat, Inc.

Westat, Inc. (1995). *Report on data collection activities for the 1994 National Assessment of Educational Progress*. Rockville, MD: Author.

Williams, P. L., Lazer, S., Reese, C. M., & Carr, P. (1995). *NAEP 1994 U.S. history: A first look—Findings from the National Assessment of Educational Progress*. Washington, DC: National Center for Education Statistics.

Williams, P. L., Reese, C. M., Campbell, J. R., Mazzeo, J., & Phillips, G. W. (1995). *NAEP 1994 reading: A first look—Findings from the National Assessment of Educational Progress*. Washington, DC: National Center for Education Statistics.

Williams, P. L., Reese, C. M., Lazer, S., & Shakrani, S. (1995). *NAEP 1994 geography: A first look—Findings from the National Assessment of Educational Progress*. Washington, DC: National Center for Education Statistics.

Wilson, D. T., Wood, R., & Gibbons, R. (1991). *TESTFACT: Test scoring, Item Statistics, and Item Factor Analysis 386/486 Version*. Mooresville, IN: Scientific Software.

Wingersky, M., Kaplan, B. A., & Beaton, A. E. (1987). Joint estimation procedures. In A. E. Beaton, *Implementing the new design: The NAEP 1983-84 technical report* (pp. 285-292). (No. 15-TR-20) Princeton, NJ: Educational Testing Service, National Assessment of Educational Progress.

Wolter, K. M. (1985). *Introduction to variance estimation*. New York, NY: John Wiley & Sons.

Yamamoto, K. (1988). Science data analysis. In A. E. Beaton, *Expanding the new design: The NAEP 1985-86 technical report* (pp. 243-255). (No. 17-TR-20) Princeton, NJ: Educational Testing Service, National Assessment of Educational Progress.

Zieky, M. (1993). Practical questions in the use of DIF statistics. In P. W. Holland & H. Wainer (Eds.), *Differential item functioning*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Zwick, R. (1986). *Assessment of the dimensionality of NAEP year 15 reading data*. (ETS Research Report No. 86-4). Princeton, NJ: Educational Testing Service.

Zwick, R. (1987). Assessing the dimensionality of NAEP reading data. *Journal of Educational Measurement, 24*, 293-308.

Zwick, R. (1988). *Professionally scored items*. Technical memorandum.

Zwick, R., Donoghue, J. R., & Grima, A. (1993). Assessment of differential item functioning for performance tasks. *Journal of Educational Measurement, 30,* 233-251.

Zwick, R., & Grima, A. (1991). *Policy for differential item functioning (DIF) analysis in NAEP*. Technical memorandum.

Zwinderman, A. H. (1991). Logistic regression Rasch models. *Psychometrika, 56*, 589-600.

☆ U.S. GOVERNMENT PRINTING OFFICE: 1997-422-447/60540

United States
Department of Education
Washington, DC 20208-5653

Official Business
Penalty for Private Use, $300

Postage and Fees Paid
U.S. Department of Education
Permit No. G-17

Standard Mail (B)

NCES 97-897

1032

BEST COPY AVAILABLE